From: Kaushal, Sumesh
Sent: Tuesday, March 23, 2004 5:22 PM
To: STIC-Biotech/ChemLib
Subject: 09938391: Sequence and Interference Search

09/938,391 Sequence and Interference Search

Please search

SEQ ID NO:2	**PRT 230**
SEQ ID NO:4	**PRT 184**
SEQ ID NO:1	**DNA 829**
SEQ ID NO:3	**DNA 555**

thanks
S. Kaushal
AU1636, REM2.B85
Ph: 571-27-20769
Mail Box: REM2.C70

Interference Search files removed – SL

Searcher:_____	TYPE OF SEARCH:	VENDOR/COST (where applic.)
Phone:_____	NA Sequences:_____	STN:_____
Location:_____	AA Sequences:_____	DIALOG:_____
Date Picked Up:_____	Structures:_____	Questel/Orbit:_____
Date Completed: 3/31/04	Bibliographic:_____	DRLink:_____
Searcher Prep/Review:_____	Litigation:_____	Lexis/Nexis:_____
Clerical:_____	Full text:_____	Sequence Sys.:_____
Online time:_____	Patent Family:_____	WWW/Internet:_____
	Other:_____	Other (specify):_____

OM nucleic - nucleic search, using sw model

Copyright   GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: March 29, 2004, 09:11:19 ; Search time 3335.77 Seconds
                                       (without alignments)
                                       10771.549 Million cell updates/sec

Title: US-09-938-391-1
Perfect score: 829
Sequence: 1 cctggcggcagatgacat.......aataaaggaggagccaaagag 829

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 347072 seqs, 2167151699S residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
    1: gb_ba:*
    2: gb_htg:*
    3: gb_in:*
    4: gb_om:*
    5: gb_ov:*
    6: gb_pat:*
    7: gb_ph:*
    8: gb_pl:*
    9: gb_pr:*
   10: gb_ro:*
   11: gb_sts:*
   12: gb_sy:*
   13: gb_un:*
   14: gb_vi:*
   15: gb_ba:*
   16: gb_fun:*
   17: em_hum:*
   18: em_in:*
   19: em_mu:*
   20: em_om:*
   21: em_or:*
   22: em_ov:*
   23: em_pat:*
   24: em_ph:*
   25: em_pl:*
   26: em_ro:*
   27: em_sts:*
   28: em_un:*
   29: em_vi:*
   30: em_htg_hum:*
   31: em_htg_inv:*
   32: em_htg_other:*
   33: em_htg_mus:*
   34: em_htg_pln:*
   35: em_htg_rod:*
   36: em_htg_mam:*
   37: em_htg_vrt:*
   38: em_sy:*
   39: em_htgo_hum:*
   40: em_htgo_mus:*
   41: em_htgo_other:*

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	829	100.0	829	6	AX399629	AX399629 Sequence
2	829	100.0	829	6	BD188110	BD188110 Method an
3	555	66.9	555	6	AX399631	AX399631 Sequence
4	555	66.9	555	6	BD188111	BD188111 Method an
5	550.4	66.4	552	6	BD240900	BD240900 Expressio
6	504.8	60.9	2637	9	BC033715	BC033715 Homo sapi
7	503.2	60.7	1388	9	AK130835	AK130835 Homo sapi
8	503.2	60.7	5408	9	AF018082	AF018082 Homo sapi
9	503.2	60.7	5929	9	AF018081	AF018081 Homo sapi
10	501.6	60.5	2364	9	AK098216	AK098216 Homo sapi
11	496.8	59.9	3394	6	I51045	I51045 Sequence 4
12	496.8	59.9	3394	6	AX409531	AX409531 Sequence
13	496.8	59.9	3394	6	HUMCOL18AX	L22548 Human colla
14	493.2	59.5	900	6	AX370851	AX370851 Sequence
15	436.8	52.7	1615	10	MMU03714	U03714 Mus musculu
16	436.8	52.7	4437	10	MUSCOLLAG	D17546 Mus musculu
17	436.8	52.7	4802	10	BC043697	BC043697 Mus muscu
18	423	51.0	4031	6	I51044	I51044 Sequence 1
19	423	51.0	4031	10	MUSCOLPRO	L22545 Mouse alpha
20	420	50.7	564	9	AF416592	AF416592 Homo sapi
21	418.4	50.5	563	6	AR431862	AR431862 Sequence
22	418.2	50.4	2376	6	AX816106	AX816106 Sequence
23	418	50.4	546	6	AR193165	AR193165 Sequence
24	418	50.4	549	6	BD240882	BD240882 Expressio
25	418	50.4	549	6	AX100086	AX100086 Sequence
26	418	50.4	552	6	BD266741	BD266741 Proteins
27	418	50.4	552	6	AX395662	AX395662 Sequence
28	418	50.4	552	6	AX473835	AX473835 Sequence
29	418	50.4	552	9	BD081407	BD081407 Fused pro
30	418	50.4	786	9	AF282883	AF282883 Homo sapi
31	418	50.4	2376	9	AX816108	AX816108 Sequence
32	411.6	49.7	555	9	AF184060	AF184060 Homo sapi
33	410.8	49.6	534	6	AR193166	AR193166 Sequence
34	410.8	49.6	537	6	AX100092	AX100092 Sequence
35	409.4	49.4	851	10	AF189709	AF189709 Rattus no
36	398.4	48.1	650	6	AX370853	AX370853 Sequence
37	394.2	47.6	552	6	BD272245	BD272245 Anti-angi
38	377.8	45.6	624	6	AX084406	AX084406 Sequence
39	376.8	45.5	573	6	BD271283	BD271283 Compositi
40	376.8	45.5	573	6	AR183471	AR183471 Sequence
41	376.8	45.5	573	6	AR183520	AR183520 Sequence
42	376.8	45.5	573	6	AR217279	AR217279 Sequence
43	376.8	45.5	573	6	AR282751	AR282751 Sequence
44	376.8	45.5	573	6	AX042272	AX042272 Sequence
45	376.8	45.5	573	6	AX490666	AX490666 Sequence

ALIGNMENTS

RESULT 1
AX399629
LOCUS   AX399629   829 bp   DNA   linear   PAT 06-JUN-2002
DEFINITION   Sequence 1 from Patent BP1191036.
ACCESSION   AX399629
VERSION   AX399629.1   GI:21135409
KEYWORDS   .
SOURCE   Canis familiaris (dog)
  ORGANISM   Canis familiaris
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1
  AUTHORS   Sheppard,M.G. and Tong,X.
  TITLE   Methods and compositions for diagnosing and treating disorders
             involving angiogenesis

Pred. No. is the number of results predicted by chance to have a

JOURNAL   Patent: EP 1191036-A 1 27-MAR-2002;
           Pfizer Products Inc. (US)
FEATURES   Location/Qualifiers
     source   1..829
                     /organism="Canis familiaris"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9615"
ORIGIN

Query Match   100.0%; Score 829; DB 6; Length 829;
Best Local Similarity 100.0%; Pred. No. 1e-99;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 CCCTGGCGGGCAGATGACATCCTGGCGGCCCCCCGGCGCCCTGCTGACCCCCAGCCCTAC   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 CCCTGGCGGGCAGATGACATCCTGGCGGCCCCCCGGCGCCCTGCTGACCCCCAGCCCTAC   60

Qy   61 CCCGGGGCCCCCGCCACCACGGCTCCTACGTGCACTTCCAGCCGGCTGGTGCTGCACCTGGTGGCG  120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCCGGGGCCCCCGCCACCACGGCTCCTACGTGCACTTCCAGCCGGCTGGTGCTGCACCTGGTGGCG  120

Qy  121 CCGTCCACACCCACCACCCACCCACCCACCAGGCATCGGAGGCATCCCACGGCCGCGGGG  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CCGTCCACACCCACCACCCACCCACCCACCAGGCATCGGAGGCATCCCACGGCCGCGGGG  180

Qy  181 TGCGAGACGTGGCGACGGAGGCCCGGGCAGGCCCACCGGCTCGTCGCTGCTGGCGG  240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TGCGAGACGTGGCGACGGAGGCCCGGGCAGGCCCACCGGCTCGTCGCTGCTGGCGG  240

Qy  241 GGCAGGCTGCTGGAGCGAGGAGGCCGCGAGCTGCCCACGGCCGCCCACGGCGGGGG  300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GGCAGGCTGCTGGAGCGAGGAGGCCGCGAGCTGCCCACGGCCGCCCACGGCGGGGG  300

Qy  301 GAGAACAGCGGCGTCATGACCTCCTTCTCCAAGACCTCCTTCTCCAAGACAGGCGGGGG  360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GAGAACAGCGGCGTCATGACCTCCTTCTCCAAGACCTCCTTCTCCAAGACAGGCGGGGG  360

Qy  361 GAGGGGCGCGCCCGCCGCCAGGAGGAGCCGCGAGCTCCGGGGGGGGACCGCTTGCCTG  420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 GAGGGGCGCGCCCGCCGCCAGGAGGAGCCGCGAGCTCCGGGGGGGGACCGCTTGCCTG  420

Qy  421 GGCAGGCTGCTGGAGCGAGGAGGCCGCCACGCCCACGCCGCCCGGGGGCCTTGCATC  480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GGCAGGCTGCTGGAGCGAGGAGGCCGCCACGCCCACGCCGCCCGGGGGCCTTGCATC  480

Qy  481 CCCCGGAAGAGCGGTGGGCACGGCTCCGACGGCTCCGACGGCCTGACCACGGCACAGCTAC  540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 CCCCGGAAGAGCGGTGGGCACGGCTCCGACGGCTCCGACGGCCTGACCACGGCACAGCTAC  540

Qy  541 TGCGAGACGTGGCGACGGAGGCCCGGGCAGGCCCACCGGCTCGTCGCTCGCTGGCGG  600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TGCGAGACGTGGCGACGGAGGCCCGGGCAGGCCCACCGGCTCGTCGCTCGCTGGCGG  600

Qy  601 GGCAGGCTGCTGGAGCGAGGAGGCCGCGAGCTGCCCACGGCCGCCCACGGTGCTCTGCATC  660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 GGCAGGCTGCTGGAGCGAGGAGGCCGCGAGCTGCCCACGGCCGCCCACGGTGCTCTGCATC  660

Qy  661 GAGAACAGCGTCATGACCTCCTTCTCCAACTAGGCCGCCCACGCAGGCGCCAGGCGGGGG  720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 GAGAACAGCGTCATGACCTCCTTCTCCAACTAGGCCGCCCACGCAGGCGCCAGGCGGGGG  720

Qy  721 GAGGGGCGCGCCCGCCGCCAGGAGGAGCCGCCTGGCCGGGACCGCTTGCCTG  780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 GAGGGGCGCGCCCGCCGCCAGGAGGAGCCGCCTGGCCGGGACCGCTTGCCTG  780

Qy  781 CACCGTCACGTTTAATGTAATCCTCAAGAATAATAAAGGAGCCAAAGAG  829
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 CACCGTCACGTTTAATGTAATCCTCAAGAATAATAAAGGAGCCAAAGAG  829
```

RESULT 2
BD188110
LOCUS   BD188110   829 bp   DNA   linear   PAT 17-JUL-2003

DEFINITION Method and composition for diagnosing and treating diseases
            relating to angiogenesis.
ACCESSION   BD188110
VERSION   BD188110.1 GI:32997849
KEYWORDS   JP 2003000268-A/1.
SOURCE   Canis sp.
ORGANISM   Canis sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1 (bases 1 to 829)
AUTHORS   Sheppard,M.G. and Tong,X.
TITLE   Method and composition for diagnosing and treating diseases
            relating to angiogenesis
JOURNAL   Patent: JP 2003000268-A 1 07-JAN-2003;
            PFIZER PRODUCTS INC
COMMENT   OS   Canis sp. (dog)
            PN   JP 2003000268-A/1
            PD   07-JAN-2003
            PF   24-AUG-2001 JP 2001254697
            PR   25-AUG-2000 US 60/227924
            PI   MICHAEL GEORGE SHEPPARD,XIAO TONG
            PC   C12N15/09,A01K67/027,A61K31/7088,A61K35/76,A61K39/395 PC
            ,A61K39/395,A61K45/00,
            PC   A61K48/00,A61P7/00,A61P9/10,A61P15/00,A61P17/00, PC
            A61P17/02,
            PC   A61P17/06,A61P19/02,A61P19/08,A61P25/00,A61P27/02,A61P27/06,
            PC   A61P29/00,
            PC   A61P31/00,A61P35/00,A61P35/02,A61P35/04,A61P37/05,A61P43/00,
            PC   C07K14/47,
            PC   C07K16/18,C07K19/00,C12N1/15,C12N1/19,C12N1/21,C12N5/10 PC
            ,C12Q1/02,G01N33/15,
            PC   G01N33/50,C12N15/00,C12N5/00
            CC   CANINE PRO ENDOSTATIN NUCLEOTIDE SEQUENCE
            FH   Key   Location/Qualifiers
            FT   source   1..829
            FT   /organism='Canis sp. (dog)'.
FEATURES   Location/Qualifiers
     source   1..829
                     /organism="Canis sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9616"
ORIGIN

Query Match   100.0%; Score 829; DB 6; Length 829;
Best Local Similarity 100.0%; Pred. No. 1e-99;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 CCCTGGCGGGCAGATGACATCCTGGCGGCCCCCCGGCGCCCTGCTGACCCCCAGCCCTAC   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 CCCTGGCGGGCAGATGACATCCTGGCGGCCCCCCGGCGCCCTGCTGACCCCCAGCCCTAC   60

Qy   61 CCCGGGGCCCCCGCCACCACGGCTCCTACGTGCACTTCCAGCCGGCTGGTGCTGGTGGG  120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCCGGGGCCCCCGCCACCACGGCTCCTACGTGCACTTCCAGCCGGCTGGTGCTGGTGGG  120

Qy  121 CTGAACAGCCCGCGAGCCGGGCGGGCATGCCGAGGCATCCCACCACCAGGACTTCCAGTGCTTC  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CTGAACAGCCCGCGAGCCGGGCGGGCATGCCGAGGCATCCCACCACCAGGACTTCCAGTGCTTC  180

Qy  181 CTGAACAGCCTCACAGCATCGTGCGCCGCATCTGTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTG  240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CTGAACAGCCTCACAGCATCGTGCGCCGCATCTGTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTG  240

Qy  241 CAGCAGGCGGCGCCGGCGCGGGGGTCGGCCGGCCCTTCGTCTTCCGTGTCGCGCGGCTG  300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CAGCAGGCGGCGCCGGCGCGGGGGTCGGCCGGCCCTTCGTCTTCCGTGTCGCGCGGCTG  300

Qy  301 CAGGACCTCTACAGACATCGTGCGCCGCATCTGTGGGGCTCGTCAACCTC  360
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CAGGACCTCTACAGACATCGTGCGCCGCATCTGTGGGGCTCGTCAACCTC  360

Qy  361 AGGGACGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGAGGCCCAGCTG  420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AGGGACGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGAGGCCCAGCTG  420
```

Right column (RESULT 4)

```
Db   181 GTGCGCCGGCGCCGACCCGCACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   379 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCAGAGGGCCAGCTGAAGCCCGGGCCCCGCATC 438
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCAGAGGGCCAGCTGAAGCCCGGGCCCCGCATC 300

Qy   439 TTCTCTTCGACGGCAGAGATGTCCTGCAGCACCCGCCTGGCCCCGGAGAGCGTGTGG 498
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TTCTCTTCGACGGCAGAGATGTCCTGCAGCACCCGCCTGGCCCCGGAGAGCGTGTGG 360

Qy   499 CACGGGCTCCGACCCCCAGCGGGCCGCCGCCACCGCAGCGTCGTCGCTGCTGGCGACG 558
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CACGGGCTCCGACCCCCAGCGGGCCGCCGCCACCGCAGCGTCGTCGCTGCTGGGGACG 420

Qy   559 GAGGCCCCGGCCACCGGGCAGGGGTCGTCGCTGCTGGGGGCAGGGCTGCTGGAGCAG 618
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAGGCCCCGGCCACCGGGCAGGGGTCGTCGCTGCTGGGGGCAGGGCTGCTGGAGCAG 480

Qy   619 GAGGCCGCGAGCTGCCGCCACCGGCCTTGCTGGTGCTCTGCATCGAGAACACAGCGTCATGACC 678
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAGGCCGCGAGCTGCCGCCACCGGCCTTGCTGGTGCTCTGCATCGAGAACACAGCGTCATGACC 540

Qy   679 TCCTTCTCCAAGTAG 693
             |||||||||||||||
Db   541 TCCTTCTCCAAGTAG 555
```

RESULT 4
BD188111
LOCUS   BD188111   555 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION Method and composition for diagnosing and treating diseases
            relating to angiogenesis.
ACCESSION   BD188111
VERSION   BD188111.1 GI:32997850
KEYWORDS   JP 2003000268-A/2.
SOURCE   Canis sp.
ORGANISM   Canis sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1 (bases 1 to 555)
AUTHORS   Sheppard,M.G. and Tong,X.
TITLE   Method and composition for diagnosing and treating diseases
            relating to angiogenesis
JOURNAL   Patent: JP 2003000268-A 2 07-JAN-2003;
            PFIZER PRODUCTS INC
COMMENT   OS   Canis sp. (dog)
            PN   JP 2003000268-A/2
            PD   07-JAN-2003
            PF   24-AUG-2001 JP 2001254697
            PR   25-AUG-2000 US   60/227924
            PI   MICHAEL GEORGE SHEPPARD,XIAO TONG
            PC   C12N15/09,A01K67/027,A61K31/7088,A61K35/76,A61K39/395 PC
            ,A61K39/395,A61K45/00,
            PC   A61K48/00,A61P17/00,A61P7/00,A61P9/00,A61P15/00,A61P17/00, PC
            A61P17/02,
            PC   A61P17/06,A61P19/02,A61P19/08,A61P25/00,A61P27/02,A61P27/06,
            PC   A61P29/00,
            PC   A61P31/00,A61P35/00,A61P35/02,A61P35/04,A61P37/06,A61P43/00,
            PC   C07K14/47,
            PC   C07K16/18,C07K19/00,C12N1/15,C12N1/15,C12N1/19,C12N1/21,C12N5/10 PC
            ,C12Q1/02,G01N3/15,
            PC   G01N33/50,C12N15/00,C12N5/00
            CC   CANINE ENDOSTATIN NUCLEOTIDE SEQUENCE
FH   Key   Location/Qualifiers
FT   source   1..555
FT   /organism='Canis sp. (dog)'.
FEATURES   Location/Qualifiers
     source   1..555
                     /organism="Canis sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9616"
ORIGIN

Left column (RESULT 3)

```
Db   361 AGGGACGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTCGGGCTCCAGAGGGCCAGCTG 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   421 AAGCCCGGGCCCCGCATCTTCTCTTCGACGGCAGAGATGTCCTGCAGCACCCGCCTGG 480
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 AAGCCCGGGCCCCGCATCTTCTCTTCGACGGCAGAGATGTCCTGCAGCACCCGCCTGG 480

Qy   481 CCCCGGAAGAGCGTGTGCACGGCTCCGACGGCTCCGACCCCCAGCGGGCCGCCGCACGCTAC 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CCCCGGAAGAGCGTGTGCACGGCTCCGACGGCTCCGACCCCCAGCGGGCCGCCGCACAGCTAC 540

Qy   541 TGGGAGACGTGGCGGACGGAGAGCCCGGCACCCGGCCAGGCGTCGTCGCTGCTGGCCG 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TGGGAGACGTGGCGGACGGAGAGCCCGGCACCCGGCCAGGCGTCGTCGCTGCTGGGCG 600

Qy   601 GGCAGGCTGCTGGAGCAGGAGGCCGCGAGCTGCCGCCTTGCTCGTGGTGCTCTGCATC 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GGCAGGCTGCTGGAGCAGGAGGCCGCGAGCTGCCGCCTTGCTCGTGGTGCTCTGCATC 660

Qy   661 GAGAACAGCGTCATGACCTCCTTCTCCAAGTAGGGCCCCCGGACCAGCAGGCGGGG 720
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 GAGAACAGCGTCATGACCTCCTTCTCCAAGTAGGGCCCCCGGACCAGCAGGCGGGG 720

Qy   721 GAGGGGGCGCCCCCAGGAGCATCCGCCCGCCCGGGGGGGCTGGCCTGGCTGCCTG 780
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GAGGGGGCGCCCCCAGGAGCATCCGCCCGCCCGGGGGGGCTGGCCTGGCTGCCTG 780

Qy   781 CACCGTCACGTTTAATGTAATCCTCAAGAAATAAAAGGAAGCCAAAGAG 829
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 CACCGTCACGTTTAATGTAATCCTCAAGAAATAAAAGGAAGCCAAAGAG 829
```

RESULT 3
AX399631
LOCUS   AX399631   555 bp   DNA   linear   PAT 06-JUN-2002
DEFINITION Sequence 3 from Patent EP1191036.
ACCESSION   AX399631
VERSION   AX399631.1 GI:21335410
KEYWORDS   .
SOURCE   Canis familiaris (dog)
ORGANISM   Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1
AUTHORS   Sheppard,M.G. and Tong,X.
TITLE   Methods and compositions for diagnosing and treating disorders
            involving angiogenesis
JOURNAL   Patent: EP 1191036-A 3 27-MAR-2002;
            Pfizer Products Inc. (US)
FEATURES   Location/Qualifiers
     source   1..555
                     /organism="Canis familiaris"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9615"
ORIGIN

Query Match   66.9%; Score 555; DB 6; Length 555;
Best Local Similarity   100.0%; Pred. No. 1e-63;
Matches 555; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy   139 CACACCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGGACCAGCCCGACAGCCCGACACAGCCCGCAGCCG 198
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CACACCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGGTGCCCTGGACACAGCCCGACACAGCCCGCAGCCCG 60

Qy   199 GGGGGCATGCGAGGCATCCGGGAGGCGGACTTCCAGCTGTTCCAGCAGGCGGCGCGCGCG 258
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GGGGGCATGCGAGGCATCCGGGAGGCGGACTTCCAGCTGTTCCAGCAGGCGGCGCGCGCG 120

Qy   259 GGGCTGGCCGCGCACCTTCCGGCCTTCCTCGTCGTCGTCGTCGTCGTCGACACCTTCACAGCATC 318
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGGCTGGCCGCGCACCTTCCGGCCTTCCTCGTCGTCGTCGTCGTCGTCGACACCTTCACAGCATC 180

Qy   319 GTGCGCCGGCGCCGACCCGCACCGGGGTGCCCGTCGTCAACCTCAGGGGACGAGGTGCTCTTC 378
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
FH   Key             Location/Qualifiers
FT   CDS             (1)..(552).
                     Location/Qualifiers
                     1..552
FEATURES             Location/Qualifiers
     source          1..552
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9615"

ORIGIN

Query Match  66.4%;  Score 550.4;  DB 6;  Length 552;
Best Local Similarity  99.8%;  Pred. No. 4.1e-63;
Matches  551;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy  139 CACACCCACCAGAGACTTCCAGTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 198
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 CACACCCACCAGAGACTTCCAGTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 60

Qy  199 GGGCGCATGCGGAGGCATCCGGGAGCGGAGACTTCCAGTGCTTCCAGCAGGCGCGCCGCG 258
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGGCGCATGCGGAGGCATCCGGGAGCGGAGACTTCCAGTGCTTCCAGCAGGCGCGCCGCG 120

Qy  259 GGGCTGGCCGGCCACCTTCCGGGCCTTCCGGGCGCGGGACCTGTCGCAGGACCTCTACAGCATC 318
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GGGCTGGCCGGCCACCTTCCGGGCCTTCCGGGCTCGCGGGACCTCACAGCATC 180

Qy  319 GTGCGCCGGCCGACCGACCGGGGTGCCGGCCGGTGCAACCTCAGGGACGACGAGGTGCTCTTC 378
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GTGCGCCGGCCGACCGACCGGGGTGCCGGTCGTCAACCTCAGGGACGACGAGGTGCTCTTC 240

Qy  379 CCCAGCTGGGAGGCCTTATTCTCGGGGTCCGAGAGCCCGCCCGCATC 438
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CCCAGCTGGGAGGCCTTATTCTCGGGGTCCGAGAGCCCGCCCGCATC 300

Qy  439 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCCCCGGAAGAGCGTGTGG 498
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCCCCGGAAGAGCGTGTGG 360

Qy  499 CACGGCTCCGACCCCAGCGGGCCGCCGCTACTGCGACCGTGGCGGACG 558
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 CACGGCTCCGACCCCAGCGGGCCGCCGCTACTGCGACCGTGGCGGACG 420

Qy  559 GAGGCCCCGGCCGGCACCGGGCCAGGCGTCGTGCTGCTCGGCGGCCAGGCCGCTGGAGCAG 618
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GAGGCCCCGGCCGGCACCGGGCCAGGCGTCGTGCTGCTCGGCGGCCAGGCCGCTGGAGCAG 480

Qy  619 GAGGCCGCGAGCTGCCGCCACGCCTTCGTGGTGCTCTGCATCGAGAACACAGCGTCATGACC 678
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 GAGGCCGCGAGCTGCCGCCACGCCTTCGTGGTGCTCTGCATCGAGAACACAGCGTCATGACC 540

Qy  679 TCCTTCTCCAAG 690
        ||||||||||||
Db  541 TCCTTCTCCAAG 552
```

RESULT 6
BC033715
LOCUS   BC033715   2637 bp   mRNA   linear   PRI 08-JUL-2002
DEFINITION Homo sapiens, Similar to collagen, type XVIII, alpha 1, clone
            IMAGE:4425380, mRNA, partial cds.
ACCESSION   BC033715
VERSION   BC033715.1 GI:21708148
KEYWORDS
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2637)
AUTHORS   Strausberg,R.
TITLE   Direct Submission
JOURNAL   Submitted (02-JUL-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA

```
Query Match  66.9%;  Score 555;  DB 6;  Length 555;
Best Local Similarity  100.0%;  Pred. No. 1e-63;
Matches  555;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  139 CACACCCACCAGAGACTTCCAGTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 198
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 CACACCCACCAGAGACTTCCAGTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 60

Qy  199 GGGCGCATGCGGAGGCATCCGGGAGCGGAGACTTCCAGTGCTTCCAGCAGGCGCGCCGCG 258
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGGCGCATGCGGAGGCATCCGGGAGCGGAGACTTCCAGTGCTTCCAGCAGGCGCGCCGCG 120

Qy  259 GGGCTGGCCGGCCACCTTCCGGGCCTTCCGGGCGCGGGACCTGTCGCAGGACCTCTACAGCATC 318
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GGGCTGGCCGGCCACCTTCCGGGCCTTCCGGGCTCGCGGGACCTCACAGCATC 180

Qy  319 GTGCGCCGGCCGACCGACCGGGGTGCCGGCCGGTGCAACCTCAGGGACGACGAGGTGCTCTTC 378
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GTGCGCCGGCCGACCGACCGGGGTGCCGGTCGTCAACCTCAGGGACGACGAGGTGCTCTTC 240

Qy  379 CCCAGCTGGGAGGCCTTATTCTCGGGGTCCGAGAGCCCGCCCGCATC 438
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CCCAGCTGGGAGGCCTTATTCTCGGGGTCCGAGAGCCCGCCCGCATC 300

Qy  439 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCCCCGGAAGAGCGTGTGG 498
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCCCCGGAAGAGCGTGTGG 360

Qy  499 CACGGCTCCGACCCCAGCGGGCCGCCGCTACTGCGACCGTGGCGGACG 558
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 CACGGCTCCGACCCCAGCGGGCCGCCGCTACTGCGACCGTGGCGGACG 420

Qy  559 GAGGCCCCGGCCGGCACCGGGCCAGGCGTCGTGCTGCTCGGCGGCCAGGCCGCTGGAGCAG 618
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GAGGCCCCGGCCGGCACCGGGCCAGGCGTCGTGCTGCTCGGCGGCCAGGCCGCTGGAGCAG 480

Qy  619 GAGGCCGCGAGCTGCCGCCACGCCTTCGTGGTGCTCTGCATCGAGAACACAGCGTCATGACC 678
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 GAGGCCGCGAGCTGCCGCCACGCCTTCGTGGTGCTCTGCATCGAGAACACAGCGTCATGACC 540

Qy  679 TCCTTCTCCAAGTAG 693
        |||||||||||||||
Db  541 TCCTTCTCCAAGTAG 555
```

RESULT 5
BD240900
LOCUS   BD240900   552 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION Expression and export of angiogenesis inhibitors as immunofusins.
ACCESSION   BD240900
VERSION   BD240900.1 GI:33050670
KEYWORDS   JP 200523036-A/20.
SOURCE   Canis familiaris (dog)
ORGANISM   Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1 (bases 1 to 552)
AUTHORS   Lo,K.M. Li,Y. and Gillies,S.D.
TITLE   Expression and export of angiogenesis inhibitors as immunofusins
JOURNAL   Patent: JP 200523036-A 20 30-JUL-2002;
            LEXIGEN PHARMACEUTICALS CORP
OS   Canis familiaris (dog)
PN   JP 200523036-A/20
PD   30-JUL-2002
PF   25-AUG-1999 JP 200056305
PR   25-AUG-1998 US 60/097883
PI   KIN MING LO, YUE LI, STEPHEN D GILLIES
PC   C12N15/09,A61K31/711,A61K38/00,A61K39/395,A61K48/00,A61P7/00,
PC   A61P35/00,
PC   A61P43/00,C07K14/745,C07K14/78,C07K19/00,C12N5/10,C12P21/02,
PC   C12N15/00,
PC   C12N5/00,A61K37/02
CC   Endostatin

REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: National Institutes of Health Intramural
             Sequencing Center (NISC),
             Gaithersburg, Maryland;
             Web site: http://www.nisc.nih.gov/
             Contact: nisc_mgc@nhgri.nih.gov
             Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
             Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
             Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
             Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
             Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
             McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
             Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
             Young,A., Zhang,L.-H. and Green,E.D.

             Clone distribution: MGC clone distribution information can be found
             through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
             Series: IRAK Plate: 68 Row: e Column: 16
             This clone was selected for full length sequencing because it
             passed the following selection criteria: Hexamer frequency ORF
             analysis.
FEATURES   Location/Qualifiers
     source   1..2637
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4425380"
                     /tissue_type="Kidney, hypernephroma"
                     /clone_lib="NIH_MGC_89"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS   <1..2453
                     /codon_start=3
                     /product="Similar to collagen, type XVIII, alpha 1"
                     /protein_id="AAH33715.1"
                     /db_xref="GI:21708149"
                     /translation="TRPPGPGREGPPGFPGLPGPPGPGPGRGPPGRTGQKGSLGEAGAP
                     GHKGSKGAPGPAGARGESGLAGAPGAGPPGAPGPPGGPLPAGPFDMEGSGGPFW
                     STARSADGPQGPPGLPGLKGDPGVPGLPGKAGEVGADGLPGFPGLPGRBGIAGPQGPK
                     GDRGSRGEKGDPGKDGDGVPGQGLPGPPGPPGPVVVYSEQDGSVLSVPGPBGRPGAGFP
                     GPAGPKGNLGSKGERGSPGPKEKGEPGSIFSPDGGALGPAQKGAKGEPGFRGPPGPY
                     GRPGYXGEIGFPGRPGPMGPMGNGLKGEKGEPGDASLGPGMRGMPGPPGPPGPPGTP
                     VYDSNVFAESSRPGPPGLPGNQGPPGPKGAKGEVGPPGPPGQFPFDFLQLEAEMKGEK
                     GDRGDAGQKGERGEPGGGGFFGSSLPGPPGPPGPPGPHRQTISVPGPPGPGPGTMGASSGVR
                     LWATRQAMLGQVHEVPEGWLIFVABQ2ELYVRVQNGFRKVQLEARTPLPRGTDNEVAA
                     LQPPVVQLHDSNPYPRREHHPTARPWRADDILASPPRLPBPQYPGAPHHSSYVHLR
                     PARPTSPPAAHSHRDFQPVLHIVALNSPLSGMRGIRGADFQCFQQARAVGLAGTFRAF
                     LSSRLQDLYSIVRRADRAAVPIVNLKDELLFPSWEALFSSSEGPKPAGRRIFSPDGKD
                     VLRHPTWPQKSVWHGSDPNGRRLTEBSYCET WRTEAPSATGQASSLLGGRLLGQSAASC
                     HHAYIVLCIENSFMTASK"

Qy   1 CCCTGCGCGGCAGATGACATCCTGGCCCGCCCCCCGCCCGGCCCGCTCTGGACACCCCTAC 60
          ||||||||||||||||||||||||||||||| ||||||||| ||||| ||||||||||||||
Db   1776 CCCTGCGCGGCAGATGACATCCTGGCCAGCCCCCTCGCCTCGCCGAGCCCCAGCCCTAC 1835

Qy   61 CCCGGGGCCCCGCACCACGGCTCCTACGTGCACTTCAGCCGCCGGCTCGCCCCACTGGTGGG 120
          ||||||||||||||||||||||||||||| ||||||||||||||||||| | |||||||||
Db   1836 CCCGGAGCCCCGCACCACACGTCCTTACGTGCACTTCAGCCGCCGGCTCGCGCCGACCCACCAAG---- 1891

Qy   121 CCCGTCCACACCCACCACCACCCACCCACCAGGACTTCCAGCTGCGTGCTGCACCTGGGCC 180
          ||||||||||||||||||||||||||||||||| ||||||||||||| ||||||||||
Db   1892 -------CCCACCCGCCCACAGCCACCGCGACTTCCAGCCGCCGGTGCTCCCACCTGGTTGCG 1943

Qy   181 CTGAACACGCCCGCAGCCCGGCGGCGGCATCGAGGCATCCGGGGAGCGGCACTTCCAGTGCTTC 240
          | |||||||||||||| |||||||||||||||||| |||| ||| ||||||| ||||||||||
Db   1944 CTCAACAGCCCCCTGTCGGCGGCGCATCGGGGGCATCGGGGCATCCGGGGCCGGACTTCCAGTGCTTC 2003

Qy   241 CAGCAGGGCGCGCGCCGGCGAGGGTGGCGGCCGCACCTTCCTTCGGCACCTTCCGGGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||
Db   2004 CAGCAGGGCGCGGCCGGGGCCGTGGGGGCGTGGGCGCCGGCACGGCTTCCGGTCTTCCGGGCTG 2063

Qy   301 CAGGACCCTCACAGCACATCGTGCCGCGCCCGACCGCACCCGGGGTGCCGTGTCGTCAACCTC 360
          |||||||||||||||| |||||||||||| ||||| ||||||||||||||||||||||||
Db   2064 CAGGACCCTACACAGCACATCGTGCCCGTGCGGACCGCGCCAGCCGTGCCCATGCTCAACCTC 2123

Qy   361 AGGGACGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTCGGCACTGGGCTCCGAGGCCAGCTG 420
          |||| ||||||||||||||||||||||||||| ||||||||||| ||||| ||||||||||||||
Db   2124 AAGGACGAGCTGCTGTTCCCCAGCTGGGAGGCTCTGTTCTCAGGCTCTGAGGGTCCCAAG 2183

Qy   421 AAGCCCGGGGCCCGCACTTCTCTTCGACGGCAGAGATGCTGCTGCAGAGATGCTCCAGAGATGCTTCGGG 480
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   2184 AAGCCCGGGGCGCAGGACACCATCTTCTTTGACGGCAGGCTCAAGGACGTCCTGAGGCACCACTGG 2243

Qy   481 CCCCGGAAGAGCGTGTGGCACGGCTCCGACCCCAGCGGCGGCCGCCTGACCGACCAGCTAC 540
          |||||||||||| ||| || || || |||||||||||||||||||||||||||||||||||
Db   2244 CCCCAGAAGAGCGTGTGGCATGGCTCGGACCCCAACGGGCGCCGCCTGACCGAGAGCTAC 2303

Qy   541 TGCGAGACGTGGCGGACCGGAGGCCCCCGCCGGCCAGGGCCAGGCCTCCCTCCGTGCTGTGGCG 600
          ||||||||| || || ||||||| ||||||||||| ||||||||||||||||||||||| |||
Db   2304 TGTGAGACGTGGGGCGGACCAGGAGGCTCCCCGCCAGGGCCAGGCCTGCCTCCTGCTGGTGGGG 2363

Qy   601 GGCAGGCTGCTGGAGCAGGAGGGCCGAGCTGCGGCCACGCCCTTCGTGGTGCTCTGCATC 660
          |||||| ||||||||||||||||||||||||||| ||||||||||||| |||||||||||
Db   2364 GGCAGGCGCTCCTGGGGCAGAGTGCCGGAGCTGCCCATCACGCGCCTACATCGTCGTCTCTGCATT 2423

Qy   661 GAGAACAGCGTCATGACCTCCTTCTCCAAGTAGGGCCCCACGGACAGGCGGGG 720
          ||||||||| | ||||| || |||||||||||||||||||||||| ||||| ||
Db   2424 GAGAACAGTTCATGACTGCCTCCAAGTAGCCACCGGCTGGATGCGGATGGGCGGGAGAGG 2483

Qy   721 GAGGGGGGCCCCCGCCAGGAGCATCCGCGCCCGGGG---------- 756
          ||||| ||||| || ||||||||||| || ||||||
Db   2484 ACCGGCGGCTCGGAGGAGAAGCCCCACCGGTGGGGCAGCAGCGCCCCAGCCCTGCCCC 2543

Qy   757 --GGCCTGCCCGGGACCGTTGCTGCACCG-TCACGTTTAATGTAATCCTCAAGAAATA 813
             || |||||| || ||||||||||| ||||||| |||||||||||||||||||||||
Db   2544 CAGGACCTGCCTGCTGCCATACTTTCCTGTATAGTTCACGTTCATGTAATCCTCAAGAAATA 2603

Qy   814 AAAGGAAGCCAAAGAG 829
          |||||||||||||||
Db   2604 AAAGGAAGCCAAAGAG 2619

```
Qy    721  GAGGGGGCGCCCGCAGGAGCATCCGCCGCCCGCCCCGGG-------------------  756
           |||||||||||||||||||||||||||||||||||||
Db   1253  ACCGGCGGCTCGGAGGAAGCCCCCACCGTGGGCAGGGAGGCGGCGGAGCGCCCTGGCCC  1312

Qy    757  --GGGCCTGGCGCGACGCTGCCTGCGCACCG-TCACGTTAATGTAATCCTCAAGAAATA  813
             |||||||||||||||||||||||||||||  |||||||| |||||||||||||||||
Db   1313  CAGGACCTGGCTGCCATACTTCCGTATAGTTCACGTTCATGTAATCCTCAAGAAATA  1372

Qy    814  AAAGGAAGCCAAAGAG  829
           |||||||| |||||||
Db   1373  AAAGGAGGCCAAAGAG  1388
```

RESULT 8
AF018082
```
LOCUS       AF018082                5408 bp    mRNA    linear   PRI 18-MAR-1998
DEFINITION  Homo sapiens type XVIII collagen (COL18A1) mRNA, alternatively
            spliced, short form, complete cds.
ACCESSION   AF018082
VERSION     AF018082.1  GI:2920536
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5408)
  AUTHORS   Saarela,J., Ylikarppa,R., Rehn,M., Purmonen,S. and Pihlajaniemi,T.
  TITLE     Complete primary structure of two variant forms of human type XVIII
            collagen and tissue-specific differences in the expression of the
            corresponding transcripts
  JOURNAL   Matrix Biol. 16 (6), 319-328 (1998)
  MEDLINE   98164096
   PUBMED   9503365
REFERENCE   2  (bases 1 to 5408)
  AUTHORS   Saarela,J., Ylikarppa,R., Rehn,M., Purmonen,S. and Pihlajaniemi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-1997) Dept. of Medical Biochemistry, University
            of Oulu, Kajaanintie 52 A, Oulu 90220, Finland
FEATURES             Location/Qualifiers
     source          1..5408
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     gene            1..5408
                     /gene="COL18A1"
     CDS             60..4070
                     /gene="COL18A1"
                     /note="alternatively spliced; short form (NC1-303)"
                     /codon_start=1
                     /product="type XVIII collagen"
                     /protein_id="AAC39659.1"
                     /db_xref="GI:2920537"
                     /translation="MAPRCPWPWPRRRLLDVLAPLVLLLGVRAASAEPERISEEVGL
                     LQLLGDPPPQQVVTQTDDPDVGLAYVFGPDANSGQVARYHPPSLFFRDFSLLFHIRPAT
                     EGPGVLFAITDSAQMVLLGVKLSGVGDHQDISLLYTEPGAGQTHTARASFRLPAFVG
                     QWTHLALSVAGGFVALYVDCEEFQRMPLARSSRGLELEPGAGLFVAQAGADPDKFQG
                     VIAELKVRRDQVSPMHCLDEEGDDSDGAFGDSSGSGLGDARRLREETGAALKPRLPA
                     PPPVTTPLAGGSSTRDSRSEVERQTTVASLGAQTLPGSDSVSTWDGSVRTPGGRVK
                     EGGLKGQKGEPGVPGPPGRAGPPGGSPCLPGPPGDTGQGFPGGTPGDVGPKGDPGVGERGPPGP
                     GPPGRDGTPGRDGEPGDPGRDGEPGDPGQGFPGPPGDTGQGFPGGTPGDVGPKGDPGVGERGPPGP
                     QGPPGPGPSFRHDKLTFIDMEGSGFGDLEALRGPRGFPGPPGPPGREGPPGRTGQKGSLGAGAPGH
                     KGSKGAPGPRAGARGSGLAGAPGPAVGPGLPGAKGEVGADGIPGPGLPGREGIAGPQGPKGD
                     RGSRGBKGDPGKDGVGQPGLPGPPGPPGPVVYSEQDGSVLSVPGPEGRPGPAGFPGP
                     AGPKGNLGSKGERGSPGPKGEGKEPGSIFSPDGALGPAQKGAKGEPGPGPGPGPTPVY
                     PGYKGEIGFPGRPGRPGMNGLKGEKGEPGDASLGFGMRGMPGPPGPPGPGPGGPPGTPVY
                     DSNVFAESSRPGPPGLPGNGQPPGPKGAKGEVGPPGPFPPDFLQLEAEMKGEKGD
                     RGDAGQKGERGEPGGGGFFGSSLPGPGRGPGRGYPGLIPGGKGSIRQGPPGPQGP
                     PGIGYESGRQGFPGPPGPPGPPGPPSFRGPHRQTISVPGPPGPPGGPPGGTMGASSGVRLW
                     ATRQAMLGQVHESVPEGWLIFVAEQELIVRVQNGFRKVQLEARTPLPRGTDNEVAALQ
                     PPVVQLHDSNYPRREHPHFTARPWRADDILASPPRLPEPQPYPGAPHHSSYVHLRPA
```

```
FEATURES             Location/Qualifiers
     source          1..1388
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="TMS08226"
                     /tissue_type="thymus"
                     /clone_lib="TMS"
                     /note="cloning vector: pME18SFL3"
ORIGIN

Query Match      60.7%;  Score 503.2;  DB 9;  Length 1388;
Best Local Similarity   77.6%;  Pred. No. 4.8e-57;
Matches 664;  Conservative 0;  Mismatches 153;  Indels 39;  Gaps 3;

Qy      1  CCCTGGCGGCAGATGACATCCTGGCCGGCCCCCGGCGCCCTGCTGACCCCCAGCCCTAC   60
           |||||||||||||||||||||| | |||||||  ||||  || |||||||||||||||
Db    545  CCCTGGCGGGCAGATGACATCCTGGCCAGCCCCCTCGCCTGGCCCCGAGCCCCAGCCCTAC  604

Qy     61  CCCGGGGCCCCGACCACAGGCTCCTAGTGCACTTCCAGCCGGCTCGCCCCACTGGTGGG   120
           ||||||| ||||||||||||| ||||||||| | |||||||||| |||||| ||| ----
Db    605  CCCGGAGCCCCGACCACCAGCTCCTAGTGCACTTCCAGCCGGCGGCGGCGACCCACAG----  660

Qy    121  CCCGTCCACACCCCACCACCCACCACCAGGAGACTTCCAGCTGGTGGTGCTGCACCTGGCGGCC   180
           ||||  |||||||||||||||||||||||| ||||||||||||||||| ||||||||||||||
Db    661  -------CCCACCGGCCCACCAGCCACCACCGCGGACTTCCAGCCGGTTGCTCCACCTGGTTGCG  712

Qy    181  CTGAACAGCCCGCAGCCGGCGGCGGCGGCCATGCGGAGGAGCATCCGGGGCGGCATCCAGTGCTTC   240
           |||||||||||||||| |||||||||||||||||||||| ||||||||| |||||||||||||||
Db    713  CTCAACAGCCCCGTGTCAGGCGGCATGCGGGGGCATCCGCCGGCGGGCCGACTTCCAGTGCTTC  772

Qy    241  CAGGACGGCGGCGGGCCGCGGGGCTGGCCGGACCTTCCGGCCTTCCTGTCGTCGCGGCTG   300
           |||||||||||||||||||||||| ||||||||| ||||||| || |||||||||||||
Db    773  CAGGAGGGCGGCGGGCCGGTGGGGCTGGGCGGCGGCACCTTCCGGCCTTCCTGTCGTCGCGGCTG  832

Qy    301  CAGGACCTTCTACAGCATCGTCGCGCGCGGCCGCGCCGACGCGGGGTGCCCGGTCGCCAACCTC   360
           ||||||||||||||||| ||||||||||||||||||||||||||| ||||||||||||||||||
Db    833  CAGGACCTGTACAGCATCGTGCGCCGGTGCCGACCGCGTGCCATCGTGACCATCGTCAACCTC  892

Qy    361  AGGGACGAGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTCGGGGTCCGAGGGCCAGCTG   420
           ||||||||||| |||||||||||||||||||  | ||||||||||||||| ||||||||||||
Db    893  AAGGACGAGAGCTGCTGTTTCCCAGCTGGGAGGCTCTGTTCTCAGGCTCTGAGGGTCCGGCTG  952

Qy    421  AAGCCCGGGGCCCGCATCTTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCCTGG   480
           ||||||||| ||||||||||||||| |||||||||||| |||||||||||||||||||
Db    953  AAGCCCGGGGCACGCATCTTCTTCTTTCCTTTGACGGCAAGGACGTCCTGAGGACCACCCCACCTGG  1012

Qy    481  CCCCGGAAGAGCGTGTGCACCGGTCGACCGCTCCGACCCCAGCGGGCCGCCGACCAGCTAC   540
           |||||||||||||||||| |||||||||||||||||||||||||||||| |||||||||||
Db   1013  CCCCAGAAGAGCGTGCACCGGTCCATGCCGTGGGGCCGCCCAACCGGGCCGCCAGGAGCTAC  1072

Qy    541  TGCGGAGCTGCGGACGCTGCGGCGAGGAGCCCCCGACGGCGCCACCGGCGTCGTCGCTGCGG   600
           |||||||| ||||||| ||||||||||||||||| |||||||||||||||||  ||||||||
Db   1073  TGTGGAGACGTGGCGGACGGAGACCGGAGCGAGCGCCCCTCGGCCCAGGCCCTCCGCTGTCCTGGG  1132

Qy    601  GGCAGGCGTCCTGGAGCAGAGGAGCGCGGCGCCAGCTGCCGCCACGCCTTCGTGGTGCTCGCATC   660
           |||||||||||| | ||||||||||||  |||||||||||| ||||||||||||||| ||||||
Db   1133  GGCAGGCTCCTGGGCAGAGTGCCGCGGCAGAGTCCGGGCCGCCATCACGGCCTACATCGTGCTCTGCATT  1192

Qy    661  GAGAACAGCAGCCGTCATGACCTCCTTCCCAAGTAGGACCGCCGGCCCCACGGACCGGGGG   720
           |||||||| |||||||||||||  |||||| |||| |  | |||||||||||  ||||||||
Db   1193  GAGAACAGCTTCATGACTGCCTTCAAGTAGCACCCGCCTCGATGCGGGGGATGGCCGGGAGG  1252
```

```
LOCUS       AF018081                5929 bp    mRNA    linear   PRI 18-MAR-1998
DEFINITION  Homo sapiens type XVIII collagen (COL18A1) mRNA, alternatively
            spliced, long form, complete cds.
ACCESSION   AF018081
VERSION     AF018081.1  GI:2920534
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5929)
  AUTHORS   Saarela,J., Ylikarppa,R., Rehn,M., Purmonen,S. and Pihlajaniemi,T.
  TITLE     Complete primary structure of two variant forms of human type XVIII
            collagen and tissue-specific differences in the expression of the
            corresponding transcripts
  JOURNAL   Matrix Biol. 16 (6), 319-328 (1998)
  MEDLINE   98164096
   PUBMED   9503365
REFERENCE   2  (bases 1 to 5929)
  AUTHORS   Saarela,J., Ylikarppa,R., Rehn,M., Purmonen,S. and Pihlajaniemi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-1997) Dept. of Medical Biochemistry, University
            of Oulu, Kajaanintie 52 A, Oulu 90220, Finland
FEATURES             Location/Qualifiers
     source          1..5929
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     gene            1..5929
                     /gene="COL18A1"
     CDS             41..4591
                     /gene="COL18A1"
                     /note="alternatively spliced; long form (NC1-493)"
                     /codon_start=1
                     /product="type XVIII collagen"
                     /protein_id="AAC39658.1"
                     /db_xref="GI:2920535"
                     /translation="MAPYPCGCHILLLFCCLAAARANLLNLWLWFNEDTSHAATT
                     IPEPQGPLPVQPTADTTHVTPRNGSTEPATAPGSPEPPSELLEDGQDTPTSAESPDA
                     PEENIAGVGAEILNVAKGIRSFVQLWNDTVPTESLARAETLVLETPVGPLALAGPSST
                     PQENGTTLWPSRGIPSSPGAHTTEAGTLPAPTPSPPSLGRPWAPLTGPSVPPPSSERI
                     SEEVGLLQLLGDPPPQQVTQTDDPVGLAYVFGPDANSGVARYHFPSLFFRDFSLLP
                     HIRPATEGPGVLFAITDSAQAMVLLGVKLSGVQDGHQDISLLYTEPGAGQTHTAASFR
                     LPAFVGQWTHLALSVAGGFVALYVDCEEPQRMPLARSSRGLELEPGAGLFVAQAGGAD
                     PDKFQGVIAELKVRRDPQVSPMHCLDEEGDSSDGAFDGSGSGLGDARELLREETGAAL
                     KPRLPAPPVTTPPLAGSSSTEDSRSEEVEQTTVASLGAQTLPGSDSVSTWDGSVRT
                     PGGRVKEGLKGQKGEPGVPGPGRAGPPGSPCLPGPPGLPCPVSPLGPAGPALQTVP
                     GQGFPGPPGRDGTPGRDGEPGDTGPGEDGPKGPGDTGPGVPKGDKGDPGVGE
                     RGPPGDGPPGPPGPSFRHDKLTFIDMEGSGFGGDLEALRGRRGFPGPGPPGVPGLP
                     GEPGRFGVNSSDVPGPAGLPGVPGREGPPGPPGLPGPGPPGEREGPPGPGREGPPGTGQKGSLGE
                     AGAPGHKGSKGAPGPAGARGESGLAGAGPAGPGPGPGPGPGPAGPDDMEGSG
                     GPFWSTARSADGPQGGPPGLPGLKGDPGVPGLPGAKGEVGADGIPGFPGLPGREGIAGP
                     QGPKGDRGSRGEKGDPGKDGVGQPGLPGPPGPPVVYSEQDGVLSVPGPEGRPGF
                     AGFPGPAGPKGNLGKGSERGSPGPGKEKGEPGSIFSPDDGALGPAGKGAKGEPGPGP
                     PGPYGRPGYKGEIGFPGRPGRPGMNGLKGEKGEPGDASLGFMRGMPGPGPGPGPGP
                     PGTPVYDSNVFAESSRPGPPGLPGNQGPPGPKGAKGEVGPPGPPGQFPFDFLQLEAEM
                     KGEKGPGRDAGQGERGEPGQFFGSSLPGPPGPPGPRGYPGIPGKGSIRGQPGP
                     PGPQGPPGIGYBGRQGPPGPGPPSFPGPHRQTISVPGPPGPPGPPGPPGTMGAS
                     SGVRLWATRQAMLGQVHEVPBGWLIFVAEQBELVYRVQNGFRKVQLEARTPLPRGTDN
                     EVAALQPPVVQLHDSNPYPRREHPPTARPWRADDILASPPRLPEPQPYPGAPHSSY
                     VHLRPARPTSPPAHSHRDFQPVLHLVALNSPLSGGMRGIRGADPQCFQQARAVGLAGT
                     FRAFLSRLQDLYSIVRRADRAAVPIVNLKDELLFPSWEALFSGSBEGPLKPGARIFSF
                     DGKDVLRHPTWPQKSVWHGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQS
                     AASCHHAYIVLCIENSFMTASK"
ORIGIN
```

Query Match   60.7%; Score 503.2; DB 9; Length 5929;
Best Local Similarity   77.6%; Pred. No. 3.1e-57;
Matches 664; Conservative 0; Mismatches 153; Indels 39; Gaps   3;

```
Qy     1 CCCTGGCGGGCAGATGACATCCTGGCCCCCCGCCCCCTGCTCGGACCCCAGCCCTAC   60
           ||||||||||||||||||||||||||||||||| |||  ||||||||||||||||||
Db  3914 CCCTGGCGGGCAGATGACATCCTGGCCCCTCGCCTGCCCGCCAGCCCTAC       3973
```

RESULT 9
AF018081

```
                     RPTSPPAHSHRDFQPVLHLVALNSPLSGGMRGIRGADPQCFQQARAVGLAGTFRAFLS
                     SRLQDLYSIVRRADRAAVPIVNLKDELLFPSWEALFSGSBEGPLKPGARIFSFPDGKDVL
                     RHPTWPQKSVWHGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHH
                     AYIVLCIENSFMTASK"
ORIGIN
```

Query Match   60.7%; Score 503.2; DB 9; Length 5408;
Best Local Similarity   77.6%; Pred. No. 3.2e-57;
Matches 664; Conservative 0; Mismatches 153; Indels 39; Gaps   3;

```
Qy     1 CCCTGGCGGGCAGATGACATCCTGGCCCCCCGCCCCCTGCTCGGACCCCAGCCCTAC   60
           |||||||||||||||||||||||||||| |||| |||  ||||||||||||||||||
Db  3393 CCCTGGCGGGCAGATGACATCCTGGCCCCTCGCCTGCCCGCCAGCCCTAC       3452

Qy    61 CCCGGGGCCCCGACCACAGGCTCCTAGTGCACTTCAGCCGGCTCGCCCCCACTCGTGGG  120
           ||||||| |||| |||||||||||| |||||||||||||||||||| ||||||
Db  3453 CCCGGAGCCCCGACCACCAGCTCCTACTGCACCTGCACCTGCGGCCGGCGACCACCACAG----  3508

Qy   121 CCCGTCCACACCCACCCACCACCAGGACTTCCAGTTGGTGCTGCACCTGGTGGCC     180
           |||||||                     |||||||||||||||||||||||||||
Db  3509 -------CCCACCGCCCCACACCGCCGGACTTCCAGCTGTCCACCTCGGTTGCG     3560

Qy   181 CTGAACACCCGCAGCCGGCGGCATGCGAGGACCAGCATCCGCGGGGCACTTCCAGTGCTTC  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3561 CTCAACAGCCCCCTGTCAGGGCGGCATCGCGGGGCATCCGCGGGGCCGACTTCCAGTGCTTC  3620

Qy   241 CAGCAGGCGCGCGCCCGCGCGGGGCTGGCCCACCTTCCTGTCGTCGCGGCTG       300
           ||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  3621 CAGCAGGCGCGGCGGCCGTGGGCGTGCGCGCTTCCGCGGCTTCCGCGGCCTG    3680

Qy   301 CAGGACCTTCTACAGCATCGTGCCGCGCGCGCCGGGGTGCCCGTCGTCAACCTC      360
           ||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  3681 CAGGACCTGTACAGCATCGTGCGGCGCGCCGACCGCGCCAGCCGTGCCCATCGTCAACCTC   3740

Qy   361 AGGGACGAGGTGCTTCTCCCCAGCTGGAGGCCTTATTCTCGGGCTCCAGCTG       420
           ||||| |||||||||||||||||||||||||||||||||||||| |||||||
Db  3741 AAGGACGAGCTGCTCTTCCCCAGCTGGGAGGCTCTGTTCTCGGGTCCGGTCCTG       3800

Qy   421 AAGCCCGGGCCCGCCATCTTCTTTGACGGCAGAGATGTCCTGCAGCACCCCGCCTGG    480
           |||||||||||||||||||| |||||||||||||||||||||||| |||||||
Db  3801 AAGCCCGGGGCACCGCATCTTCTTTGACGGCAAGGACGTCCTGAGGCACCCACCTGG     3860

Qy   481 CCCCGGGAAGAGCGTGTGGCACGGCACGCTCCGACCCCAGCCGGGCGCGCCGACCAGCTAC  540
           ||||||||||||||||||| |||||||||||||||||||||||||| |||||||
Db  3861 CCCCAGAGAGAGCGTGTGGCATGGCGTGGCCAGGCTGACCGAGAGCTAC        3920

Qy   541 TGCGAGACGTGGCGGCACCGGAGGCCCCGCCAGCGGTCGCTGCTCGGGCG         600
           ||||||||||||||||||| ||||||||||||||||||||||||| ||||||
Db  3921 TGTGAGACGTGGCGGCACCGGAGGCTCCCGCGGGCCAGGCTCGCTGCTCGTGGGG    3980

Qy   601 GGCAGGCTGCTGAGACGAGGAGGCCCGAGCTGCCGCACCGCCTTCGTGGTGCTCTGCATC  660
           ||||||||||||||| ||||||||||||||||||||||||||||||| ||||||
Db  3981 GGCAGGCTCCTGGGGCAGAGTGCGCGAGCTCCTGGCGCCATCACGTCTCTGCATT      4040

Qy   661 GAGAACAGAGCTCATGAGCCTCCTTCTCCAAGTAGGCGCCGCCACCGGACAGCGGCGGGG  720
           ||||||||| |||||| ||||||||||||||||||||||||||||||||| ||||||
Db  4041 GAGAACAGTTCATGACTGCCTCCAAGTAGCCACCGCCTGGATGCCGGATGCCCGGAGAGG   4100

Qy   721 GAGGGGGGCGCCCGCAGGAGCATCCGCGCCCGGGG----------------        756
           ||||| |||||||||||||||||||||||||||||
Db  4101 ACCGGCGGCTCGGAGGAAGCCCCACCGTGGGCAGGAGGCCGCCCAGCCCTGCCCC      4160

Qy   757 --GGGCCTGCCCGGGACCGTTGCCTGCACCG-TCACGTTTAATGTAATCCTCAAGAAATA  813
             ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db  4161 CAGGACCTGCCCATACTTTCCTGTATAGTTCATGCGTTTCATGTAATCCTCAAGAAATA     4220

Qy   814 AAAGGAAGCCAAAGAG      829
           |||||||||||||||||
Db  4221 AAAGGAAGCCAAAGAG       4236
```

REFERENCE   2 (bases 1 to 2364)
  AUTHORS   Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B.,
            Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and
            Isogai,T.
  TITLE   NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2364)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE   Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES   Location/Qualifiers
     source   1..2364
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UTERU2002964"
                     /tissue_type="uterus"
                     /clone_lib="UTRU2"
                     /note="cloning vector: pME18SFL3"
ORIGIN

Query Match   60.5%; Score 501.6; DB 9; Length 2364;
Best Local Similarity   77.5%; Pred. No. 6.7e-57;
Matches 663; Conservative   0; Mismatches 154; Indels 39; Gaps 3;

Qy   1 CCCTGGCGGCAGATGACATCCTGGCGCGGCCCCCGCGGCCCCGCGGCCCTGCTGCAGCCCTAC   60
          ||||| |||||||||| ||| ||||||||| |||||||| ||||||||||||| ||||||||
Db   1518 CCCTGGCGGCAGATGACATCCTGGCGCGGCCCCCGCGGCCCCGCGGCCCTGCTGCAGCCCTAC 1577

Qy   61 CCCGGGGCCCCGCGACCACGGCTCCTACGTGCACTTCCAGCGCGCTCGCCCCACTGGTGGG 120
          |||||| ||||||||| ||||||||||| |||||||||||||||| |||||||||| ||||
Db   1578 CCCGGGGCCCCGCGACCACCAGCTCCTACGTGCACTTCCAGCGCGCTCGCGCACGACCACAAG--- 1633

Qy   121 CCCGTCCACACCCACACCCACACCCACAGACTTCCAGCGCGGCATCCCAGTGCTTC 180
          ||||||| ||||| ||||| |||||||||||||||||||||| | |||||||||
Db   1634 -------CCCACCCGCCCACACAGCCACCGCGGCGACTTCCAGCGCCGCGGGTCTCCACCTGGTTGCG 1685

Qy   181 CTGAACAGCCCAGCCTGTCAGGCGGGATGTCCTGCAGCAGATGTCCTGCAGCACCCGCGCTGG 240
          ||||||||||||| ||||||||| |||| |||||||| ||| |||||||| ||| ||||||
Db   1686 CTCAACAGCCCCGTCAGGGCGGGGCATCGGGGGCATCCCGGGCGGGGCCGGCGGGGACTTCCAGTGCTTC 1745

Qy   241 CAGCAGCAGGCGCGGCGGACGGAGGCCCCGGCCGCGGCGGCGCTCGTCGTCGCTGCTGCGG 300
          |||||||||||||||||||| ||||||||||| |||||||| ||||| ||||||||||||
Db   1746 CAGCAGCAGGCGCGGCGGACGGAGGCCCCGGCCCACCCTTCCGGTGCGTCGTCGTCGGGCTG 1805

Qy   301 CAGGAGACTTCACAGCATCGTGCGCGGCGGGCCCCGCGGGGGTGCCCCGGTGCCCATCGTCAACCTC 360
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   1806 CAGGAGACTTCACAGCATCGTGCGCGGCGGGCCCCGCGGGGGTGCCCCGGTGCCCATCGTCAACCTC 1865

Qy   361 AGGGACGAGGTGCTCTTCCCCAGCTCGGAGAGCCTTATTCTCGGGCTCCGAGGGCCCAGCTG 420
          |||||||||| |||| |||||||||||||| ||| ||| |||||||| |||||||||||||||
Db   1866 AAGGACGAGCTGCTGTTCCCAGCTGGAGGCTCTCGTTCTCAGGCTCGTGAGGGTCCGGTG 1925

Qy   421 AAGCCCGGGGCCCGCATCTCTCTTCGACGGCAGAGATGTCTGCAGCACCCGGCCTG 480
          |||||||||| ||||||||| |||||||||||| ||||||||| |||||||||||||||
Db   1926 AAGCCCGGGGCACGCATCTTTCCTTTGACGCGCAAGGACGTCCTCAGCACCCACCTG 1985

Qy   481 CCCCGGAAGAGGCGTGGCACGGCTCCGACGGCGGCGGCGGCGGCGGCCTGACCAGCTAC 540
          ||||| ||||||||||||||||||| ||||||||||||| ||||||||||||||||||||
Db   1986 CCCCAGAAGAGGCGGTGTGCATGGGCTCGGACCCACGGCGGCCCAAGCGGCGCAGCAGCTAC 2045

RESULT 10
AK098216
LOCUS   AK098216   2364 bp   mRNA   linear   PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ40897 fis, clone UTERU2002964, highly similar
            to Homo sapiens type XVIII collagen (COL18A1) mRNA.
ACCESSION   AK098216
VERSION   AK098216.1 GI:21758185
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
            Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,

Qy   61 CCCGGGGCCCCGCGACCACGGCTCCTACGTGCACTTCCAGCGCGCTCGCCCCACTGGTGGG 120
          |||||| ||||||||| ||||||||||| |||||||||||||||| |||||||| |||----- 4029
Db   3974 CCCGGAGCCCCGCACCACAGCTCCTACGTGCACTTCGGGCGGCGGGGCGGGGACCCACACAAG---

Qy   121 CCCGTCCACACCCACACCCACACCCACAGGACTTCCAGCGCGGCATCCCAGTGCTTC 180
          |||||||| ||||||||||||||||||||||||||||||||||||| |||||||| 4081
Db   4030 ------CCACCCGCCCACAGCGCCACCGCGGCGACTTCCAGCGCGTGCTTCCAGTGCTTCGG

Qy   181 CTGAACAGCCCAGCCTGTCAGGCGGGCATGCGGAGCGCATCCAGCGCGGCGGACTTCCAGTGCTTC 240
          |||| || |||||||||||||||||||||||||||||| |||||||||| |||| |||||||| 4141
Db   4082 CTCAACAGCCCCGTCAGGGCGGGGCATGCGGCGGCATGCGGGCGGGCCCGACTTCCAGTGCTTC

Qy   241 CAGCAGGCGCGGCGGACGGAGGCCCCGGCCGCGGCGGCGCTCGTCGTCGCGGCGCTG 300
          |||||| |||||||||||||||||||||||||| ||||||| |||||||| |||||| 4201
Db   4142 CAGCAGGCGCGGGGGCTGGCGGGCACTTCCGGCCTTCCGTCCTCGTCGTCGGCGCCTG

Qy   301 CAGGAGACTTCTACAGCATCGTGCGCCGGACCGGAGGGCGCACCGGGGTGCCCGGGGTGCCGTGCTGCAACTTC 360
          |||||||||||||||||||||||||| |||| |||||||||||||||||||| |||||||||||||| 4261
Db   4202 CAGGAGACTTCTACAGCATCGTGCCGCGGGCCCAGCCGTGCTGCAACATCGTGCAACATCCAACCTC

Qy   361 AGGGACGAGGTGCTCTTCCCCAGCTGGGAGGGCCTTATTCTCGGGCTCCGAGGGGCCCAGCTG 420
          |||||||||||||||| |||||||||||||||| ||||| |||||||| ||||||||||| 4321
Db   4262 AAGGAGGAGCTGCTGTTTCCAGCTGGGGAGGGCTCTGGTTCTCAGGCTCTGAGGGGTCCGGTG

Qy   421 AAGCCCGGGGCCCATCTTCTCTTTCGCACGGCAGAGATGTCCTGCAGCACCCGCCTGG 480
          |||||||||| |||| ||| |||||||||| |||||||||| |||||||||||||||| 4381
Db   4322 AAGCCCGGGGCACGCATCTCCTTTGACGGGCACAGCGTGGCTCCTGAGGCACCCCACCTGG

Qy   481 CCCCGGAAGAGGCGTGGCACGGCTCCGACGGCCGCGGCCTGACCAGCTAC 540
          ||||| |||||||| ||||||| ||| |||| | |||||| ||||||||| 4441
Db   4382 CCCCAGAAGAGAGCCGTGTGGCATGGCTCGGATGCTCGGACCGCTGACCGAGAGCTAC

Qy   541 TGCGAGACGTGGCCGGACGGAGGCCCCGGCCGCTGCGGCCACCGGGCGTCGTCGCTGCTGCCG 600
          ||||||||||| ||||||||| ||||||| |||||| |||||||||||||||||| |||| | 4501
Db   4442 TGTGAGACGTGGCCGGACGGAGGCCCCGGCCGCTGCGGCCAGGCGGCCAGGGGCTGCTGCTGGGG

Qy   601 GGCAGGCTGCTGTGGAGCAGGAGAGCGGCCGCGGAGCGTCGCCGCCACCGGCGCTCTCGTCATC 660
          |||||||||||||||||| |||||||||||| ||||| |||||||||||| ||||| ||||| 4561
Db   4502 GGCAGGCTCTCGGGGGGCAGAGTGCCGGAGCTGCCGGAGCTGCCTACAGCGTGTCGTGCTCTGCATT

Qy   661 GAGAACAGGCTCATGACCTCCTTCTCCAAGTAGGGCGCGGCCCACGGACAGAGGCGGGGG 720
          |||||||||||| |||||||| ||| |||||||||||||||| ||||||||||| ||||| 4621
Db   4562 GAGAACAGTTCATGACTGTCCTCCAAGTAGCCACCGCGGCGGCCTGGCGGATGGCCCGGAGAGG

Qy   721 GAGGGGGCGCGCCCGCGAGGAGCATCCGCCCCCGGG---------------- 756
          ||||| ||||||| |||||||| ||||||||||||
Db   4622 ACCGGCGGCTCGGAGGAGGGAGGCCCGCGGCCGCGGCCGCGCCCTGCCCC 4681

Qy   757 --GGGCCTGGCCGGGACGCTTGGCTGCCACCG-TCACGTTTAATGTAATCCTCAAGAAATA 813
             ||||||| |||||| ||||| |||||||||| ||||||||||| ||||| ||||||||| 4741
Db   4682 CAGGACCCTGGCCTGCGCCATACTTTCCTGTATAGTTCAGTTTCATGTAATCCTCAAGAAATA

Qy   814 AAAGGAAGCCAAAGAG 829
          |||||||||||||||| 4757
Db   4742 AAAGGAAGCCAAAGAG

```
Qy    361 AGGGACGAGGTGCTCTTCCCAGTGGAGGCCTATTCTCGGGCTCCGAGGGGCCAGCTG  420
          |||||||||||||||||||||||||||||| ||||||| | ||||||||||||||||
Db   1726 AAGGACGAGCTGCTGTTTCCCAGCTGGAGGGAGGCTCTGTTCTCAGGCTCTGAGGGGTCCGCTG 1785

Qy    421 AAGCCCGGGACCCGCATCTTCTTTCGACGGCAGAGATGTCCTGCAGCACCCGCCTGG  480
          ||||||||||||||||||| |||  ||| ||| ||| ||||| ||||||||||||||
Db   1786 AAGCCCGGGCACGCATCTTCCTTTGACGGCAAGGACGTCCTGAGCGCACCCACCTGG 1845

Qy    481 CCCCGGAAGAGCGTGGCGACGGCCGCCCAGCGGCCGCCGCTGACCGACAGCTAC  540
          ||||  ||||||||||||||||||||||||| |||||||||||| ||||||||||
Db   1846 CCCCAGAAGAGCGTGGCGATGGCTCGGACCCCAACGGGCTGACCGAGGAGCTAC 1905

Qy    541 TGCGAGACGTGGCGGACGAGGAGGCCCGGCCCGGCGGCCACCGAGGCGTCGTGCTGCTGGCG  600
          | | ||||||||| ||||| ||| |||||| ||||| |||||| ||||| ||||||||||||
Db   1906 TGTGAGACGTGGCGGACGAGGAGGCCCGGCCCGGCGGCCACGGGCCAGGCCTCCTGCTGGGG 1965

Qy    601 GGCAGGCTGCTGGAGCAGGAGGCCCGCAGTGCCGCCCAGCCTTCGTGGTGCTCGCATC  660
          ||||||||||||||||||||||||||||||||||||| ||||| |||||||||||||||
Db   1966 GGCAGGCTCCTGGGGCAGGAGTGCCCGCAGTGCCGCCACCTGCATCAGCCTACATCGGCTCTGCATT 2025

Qy    661 GAGAACAGCGTCATGACCTCCTTCTCCAAGTAGGGCCGGCCCACGGACAGGCGGGG  720
          |||||||||||||||||||||||||||||||||| ||||||||||  ||||||||||
Db   2026 GAGAAACAGCTTCATGACTGCCTCCAAGTAGCCACCGCTGATGCAGATGCCGGAGAGG 2085

Qy    721 GAGGGGCGCCCCGCAGGAGCATCCGCGCCCGGGG------------------------  756
          |||||||||||||||||||||||||||||||||
Db   2086 ACCGGGCGGCGGCGGCGTCGGAGGAAGCCCCACCGTGGGCAGGGAGCAGCCGGCCCCTGGCCC 2145

Qy    757 --GGGCCTGGCCGGGACGCGTTGCCTGCACCG-TCACGTTTAATGTTAATCCTCAAGAAATA  813
            |||||||||||||||  |||||||||  ||  ||||||  |||||||  ||||||||||||
Db   2146 CAGGACCTGGCTGCCATACTTTCCTGTATAGTTCACGTTTCATGTAATCCTCAAGAAATA 2205

Qy    814 AAAGGAAGCCAAAGAG  829
          ||||||||||||||||
Db   2206 AAAGGAAGCCAAAGAG 2221
```

RESULT 12
AX409531
LOCUS   AX409531   3394 bp   DNA   linear   PAT 14-JUN-2002
DEFINITION Sequence 2178 from Patent WO0229103.
ACCESSION   AX409531
VERSION   AX409531.1 GI:21142236
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
  TITLE   Gene expression profiles in liver cancer
  JOURNAL   Patent: WO 0229103-A 2178 11-APR-2002;
            GENE LOGIC INC (US)
FEATURES   Location/Qualifiers
     source   1..3394
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
                     /note="EMBL/GenBank Accession No. L22548"

ORIGIN

Query Match   59.9%; Score 496.8; DB 6; Length 3394;
Best Local Similarity   77.1%; Pred. No. 2.5e-56;
Matches 660; Conservative   0; Mismatches 157; Indels 39; Gaps 3;

```
Qy      1 CCCTGGCCGGCAGATGACATCCTGGCCGGCCCCCCCGCCCTGCTGGACACCCGCCAGCCTAC   60
          |||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db   1378 CCCTGGCCGGCAGATGACATCCTGGCCCAGCCCCCGAGCCCCCAGCCCCAGCCCTAC 1437

Qy     61 CCCGGGACCCCGCCACCACGGCTCCTACGTGCACTTCAGCCGTCGCCCACTGTGGG  120
          ||||||||||||||||||||||||||| |||||||||||||||||||
Db   1438 CCCGGAGCGGCAGCCCGACCACCAGCTCCTACGTGCACTGCACCAGCCCCCGCCACCACCAAG----- 1493
```

Query Match   59.9%; Score 496.8; DB 6; Length 3394;
Best Local Similarity   77.1%; Pred. No. 2.5e-56;
Matches 660; Conservative   0; Mismatches 157; Indels 39; Gaps 3;

```
Qy    541 TGCGAGACGTGGCGGACGAGGAGGCCCGGCCCGGCGGCCACCGAGGCGTCGTGCTGCTGGCG  600
Db   2046 TGTGAGACGTGGCGGACGAGGAGGCCCGGCCCGGCGGCCACGGGCCAGGCCTCCTGCTGGGG 2105

Qy    601 GGCAGGCTGCTGGAGCAGGAGGCCCGCAGTGCCGCCCAGCCTTCGTGGTGCTCGCATC  660
Db   2106 GGCAGGCTCCTGGGGCAGGAGTGCCCGCAGTGCCGCCATCACGGCTCTACATCGGCTGCTGCATT 2165

Qy    661 GAGAACAGCGTCATGACCTCCTTCTCCAAGTAGGGCCGGCCCACGGACAGGCGGGG  720
Db   2166 GAGAAACAGCTTCATGACTGCCTCCAAGTAGCCACCGCTGATGCAGATGCCGGAGAGG 2225

Qy    721 GAGGGGCGCCCCGCAGGAGCATCCGCGCCCGGGG-------------------------  756
Db   2226 ACCGGGCGGCGGCGGCGTCGGAGGAAGCCCCACCGTGGGCAGGGAGCAGCCGGCCCCTGGCCC 2285

Qy    757 --GGGCCTGGCCGGGACGCGTTGCCTGCACCG-TCACGTTTAATGTTAATCCTCAAGAAATA  813
Db   2286 CAGGACCTGGCTGCCATACTTTCCTGTATAGTTCACGTTTCATGTAATCCTCAAGAAATA 2345

Qy    814 AAAGGAAGCCAAAGAG  829
Db   2346 AAAGGAAGCCAAAGAG 2361
```

RESULT 11
I51045
LOCUS   I51045   3394 bp   DNA   linear   PAT 07-OCT-1997
DEFINITION Sequence 4 from patent US 5643783.
ACCESSION   I51045
VERSION   I51045.1 GI:2472748
KEYWORDS   .
SOURCE   Unknown.
  ORGANISM Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 3394)
  AUTHORS   Olsen,B.R. and Oh,S.P.
  TITLE   Collagen and uses therefor
  JOURNAL   Patent: US 5643783-A 4 01-JUL-1997;
FEATURES   Location/Qualifiers
     source   1..3394
                     /organism="unknown"
                     /mol_type="unassigned DNA"

ORIGIN

Query Match   59.9%; Score 496.8; DB 6; Length 3394;
Best Local Similarity   77.1%; Pred. No. 2.5e-56;
Matches 660; Conservative   0; Mismatches 157; Indels 39; Gaps 3;

```
Qy      1 CCCTGGCCGGCAGATGACATCCTGGCCGGCCCCCCCGCCCTGCTGGACACCCGCCAGCCTAC   60
Db   1378 CCCTGGCCGGCAGATGACATCCTGGCCCAGCCCCCGAGCCCCCAGCCCCAGCCCTAC 1437

Qy     61 CCCGGGACCCCGCCACCACGGCTCCTACGTGCACTTCAGCCGTCGCCCACTGTGGG  120
Db   1438 CCCGGAGCGGCAGCCCGACCACCAGCTCCTACGTGCACTGCACCAGCCCCCGCCACCACCAAG---- 1493

Qy    121 CCCGTCCACACCCACCACCACCACCGCCCAGCACCCACCGCCACCACCGGTTGTGTCGACCTGGTGGCC  180
Db   1494 -------CCCACCCGCCCACACCCACCACCACCGCCCAGCACCCACCGGTGTCCACCTGGTTGCG 1545

Qy    181 CTGAAACAGCCCGCCGGCCGGGACGCCGGGCATGCGAGGCAGCATCCGGGGAGCCGACTTCCAGTGCTTC  240
Db   1546 CTCAACAGCCCCGTCAGGCGGCATGGGGGCATGCGGGGCATCCGCGGGGCCGACTTCCCAGTGCTTC 1605

Qy    241 CAGCAGGCGCGCGCCCCGCGGGGCTGGGCTGCGGCGGCCACCTTCCGCGCCTTCTCGTCCTCGGCGGCTG  300
Db   1606 CAGCAGGCGCGGCGGCCCGTGGGCGTGGGCTGCGGCGGCCACCTTCCGCGCCTTCTCGTCCTCGGCCGCTG 1665

Qy    301 CAGGACCTTCTACAGCAATCGTGCGCGCGCGACCACGCCGGGGTGCCCGTCGTCGTCAACCTC  360
Db   1666 CAGGACCTGTACAGCAATCGTGCGCGCGCGACCACAGCTCCTACGTGCACTGCACCACCGACCAAG 1725
```

COMMENT   Original source text: Homo sapiens cDNA to mRNA.
FEATURES   Location/Qualifiers
     source   1..3394
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /map="21q22.3"
                     /tissue type="liver"
                     /dev_stage="adult"
     gene   1..3394
                     /gene="COL18A1"
     CDS   <1..2055
                     /gene="COL18A1"
                     /codon_start=1
                     /product="collagen type XVIII alpha 1"
                     /protein_id="AAA51864.1"
                     /db_xref="GI:562794"
                     /translation="GEVGADGIPGFGLPGREGIAGPQGPKGDRGSRGBEKGDPGKDGL
                     GQPGLPGPGRGPPGPVVVSEQDGSVLSVPGPEGPGRRGPAGPPAGPGPKGNLGSKGELGS
                     PGPKGEKGEPGSIFSPDGGALGPGMRGMPGPPGPGPPGPPGPTPVYDSNVFABSSRPGPPG
                     PGMNGLKGEKGEPGDASLGFGMRGMWGPPGPGPFFDFLQKEAEMKGEKGDRGDAGQKGBRGEPGG
                     LPGNQGPPGPKGPKGEVGPPGPGQFPFDFLQKEAEMKGEKGDRGDAGQKGBRGEPGG
                     GGFFGSSLPGAPGAPGRGYPGIPGKGEBIRGQPGPPGPGPPGPIGYEGRGQGPPGPP
                     GPGPPSFPGPHRQTISVPGPPGPPGPPGTMGASSGQVRLWATRQAMLGQVHEVP
                     EGWLIFVABQEELYRVRVQNGFRKVQLBARTPIPRGTNEVAALQPPVVQLHDSNPYPR
                     REHPHTARPWRADDLLASPPGLPEPOYPGGPHHSSYVHCGPARPISPPAHSHRDFQ
                     PVLHLVALNSPLSGGMRGIRGADFGCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRAD
                     RAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGS
                     DPNGRRLTESYCBTWRTEAPATGQASSLLGQSAASCHHAYIVLCIENSFWTA
                     SK"

ORIGIN

Query Match   59.9%;   Score 496.8; DB 9; Length 3394;
Best Local Similarity 77.1%; Pred. No. 2.5e-56;
Matches 660; Conservative 0; Mismatches 157; Indels 39; Gaps 3;

```
Qy      1 CCCTGGCGGGCGGCAGATGACATCCTGGCCGCGCCCCCGGCCCGCTGCTGGACCCCAGCCTAC 60
          |||||||||| ||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   1378 CCCTGGCGGGCGGCAGATGACATCCTGCCGCCAGCCCCCTGGGCTGCCCCAGCCCAGCCCTAC 1437

Qy     61 CCCGGGGCCCCGCCACCACGGCTCTACGTGCACTTCCAGCCGCTCGCCCCACTGGTGGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   1438 CCCGGAGGCCCGGCCACCACAGTCCTACGTGCACTGCACTTCCAGCCGCTCGCGCGACCCACAAG---- 1493

Qy    121 CCCGTCCACACCCACCACCACCAGGACTTCCAGCCGCTGGTGCTGCACCTGGTGGCC 180
          |||||||                 |||||||||||||||||||||||||||||||
Db   1494 -------CCCACCGCCCACAGCCCACCACGGCCTGGTGCTGCTGCACCTGGTTGCG 1545

Qy    181 CTGAACAGCCGGCTGCGCGCGGCCGCGCAGGCCCCGCGGGAGGCGGACTTCCAGTGCTTC 240
          ||||||||||| |||||||||||||||||||||||||||||||||
Db   1546 CTCAACAGCCCCCTGTCAGGCGGGCATGCGGGGGCATCCGGGGCCTTCCTGTGTCCTCGGGGCTTC 1605

Qy    241 CAGCAGGCGGCCGCGGCCGGGGCTGGCGCCGCGCGCGCCGCACCGGGGTGCCCGTCGTCAACCTC 300
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1606 CAGCAGGCGGGCGGCCCGGGCGGGCGGGCTGGGCTGACCGCAGCACCTTCCGTGTCCTCGTCGGGCTG 1665

Qy    301 CAGGACCTCTACAGCATCGTGCGCGCGGCCTTATTCTCGGGAGGCCCAGCTG 360
          ||||||||||||||||||||||||||||||||||||||||
Db   1666 CAGGACCTGTACAGCATCGTGCGCGCGGCCATCCGTGCCGTGCCCATCGTCAACCTC 1725

Qy    361 AGGGACGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTCCGGGGTCCGAGGGGCCAGCTG 420
          ||||||||||||||||||||||||||||||||||||||||||||
Db   1726 AAGGACGAGCTCTGTTTCCCAGCTGGGAGGCTCTGTTCTCCAGGCTCTGAGGGTTCCGGCTG 1785

Qy    421 CCCCGGAAGAGCCCGACCATCTTCTTTGACGGCAAGGACGTCCTGCAGCACCCCGGCCTGG 480
          ||||||||||||||||||||||||||||||||||||
Db   1786 AAGCCCCGGGGCACGCATCTTCCTTTTGACGGCAAGGACGTCCTGAGGCACCCCACCTGG 1845

Qy    481 CCCCGAGAGAGCGGTGTGGCACGGAGGCGCCCCCGGCCGGCGGGCCCCAACGGCACCACGGCTC 540
          |||||||||||||||||||||||||||||||||||
Db   1846 CCCCAGAGAGACGTGTGGCATGGCGTGGGGCCCAGCTGACCGAGAGCTAC 1905

Qy    541 TGCGAGACGTGGCGACGGAGCGCCCCGGCCGGCGGGCCCCAACGGCACCACGGCTCGTCGTGGGCG 600
```

Db 499 CAGGACCTGTACAGCATCGTGCGCCGTGCCGCC...CCAGCGCGTGCCCATCGTGCCAACCTC 558
Qy 361 AGGGACGAGGTGCTCTTCCCCAGTCGGCCGGAGGCCTTATTCTCGGGCTCCGGCGCCAGCTG 420
Db 559 AAGGACGAGCAGCTGCTGTTTCCCAGCTGGGAGCGTCGTTCTCGTTTCTCAGCCTCAGGGCTG 618
Qy 421 AAGCCCGGGCCCCATCTTCTTCTTCGACGGCGAGAGATGTCCTGCAGACACCCGCCTGG 480
Db 619 AAGCCCGGGCCACGCATCTTCTTTGACGGCAAGGACGTCCTGAGGCACCCCACCTGG 678
Qy 481 CCCGGAGAGAGCGTGTGCACGGCTCCGACGCGCCGCGGGCGCCTGACCGACCACAGTAC 540
Db 679 CCCCAGAGAGAGCGTGTGCATGGCTCCGACCTCCGGACCCCCAACGGGGCCTGACCGAGAGTAC 738
Qy 541 TGCGAGACGTGCGGACGGAGGCCCCGGCGGCCACCGGGCGTCGTCGCGCTGCTGGCG 600
Db 739 TGTGAGACGTGGGGGACCGGAGGCTCCTCGGCACGGGCCCACGGGCCTCTCGCTGCTGGGG 798
Qy 601 GGCAGGCGTCCTGGAGCAGGAGGCCGGGGCCGGGCCCCACGGCCTTCGTGGTGCTCGTCATC 660
Db 799 GGCAGGGCTCCTGGGGCAGGAGGCCGGGGAGGTGCCCGCCATCACGGCTACATCGTGCTCTGCATT 858
Qy 661 GAGAACAGGGTCATGACCTCCTTC 684
Db 859 GAGAACAGCTTCATGACTGCCTCC 882

RESULT 15
MMU03714
LOCUS   MMU03714
DEFINITION Mus musculus BALB/c alpha 1 type XVIII collagen (COL18A1) mRNA,
            partial cds.
ACCESSION   U03714
VERSION   U03714.1 GI:487733
KEYWORDS   .
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (sites)
  AUTHORS   Rehn,M.V. and Pihlajaniemi,T.
  TITLE   A previously uncharacterized collagen chain, alpha 1(XVIII), with
            frequent interruptions in the collagenous sequence, a distinct
            tissue distribution, and homology with type XV collagen
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. (1994) In press
REFERENCE   2 (bases 1 to 1615)
  AUTHORS   Rehn,M., Hintikka,E. and Pihlajaniemi,T.
  TITLE   Primary structure of the alpha 1 chain of mouse type XVIII
            collagen, partial structure of the corresponding gene, and
            comparison of the alpha 1(XVIII) chain with its homologue, the
            alpha 1(XV) collagen chain
  JOURNAL   J. Biol. Chem. 269 (19), 13929-13935 (1994)
  MEDLINE   94245707
  PUBMED   8188673
REFERENCE   3 (bases 1 to 1615)
  AUTHORS   Rehn,M.V.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-NOV-1993) Marko Rehn, Department of Medical
            Biochemistry, University of Oulu, Kajaanintie 52 A Fin-90220, Oulu,
            Finland
FEATURES   Location/Qualifiers
     source   1..1615
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
                     /sex="male"
                     /tissue_type="liver"
                     /clone_lib="Mouse adult liver cDNA library by M. Rehn"
                     /dev_stage="adult, 9-10 weeks"
     CDS   <1..1449
                     /standard_name="COL18A1"

Db 1906 TGTGAGACGTGGCGGACGGAGGCTCCTCGGCACGGGCCCACGGGCCTCCTCGCTGCTGGGGG 1965
Qy 601 GGCAGGCGTCCTGGAGCAGGAGGCCGGGGCCGGGCCCCACGGCCTTCGTGGTGCTCTGCATC 660
Db 1966 GGCAGGGCTCCTGGGGCAGGAGGCTGCCGAGCTGGCCTGATCACCGCCTACATCGTGCTCTGCATT 2025
Qy 661 GAGAACAGCTCATGACCTTCTTCTTCCAAGATAGGGCGCGGCCACCGGACCAGCAGGCGGGG 720
Db 2026 GAGAACAGCTTCATGACTGCCTCCAAGTAGCCACCGCCTGGATGCAGATGGCCGGAGAGG 2085
Qy 721 GAGGGGGGCCCGCAGAGCATCCGCGCCGCCGGGG--------- 756
Db 2086 ACCGGCGGCTCGGAGAAGCCCCCACCGTGGGCAGGGAGCAGCCCGGCCCCCTGGCCC 2145
Qy 757 --GGGCCTTGGCGGGACCCTTGCCTGCACCG-TCACGTTTAATGTAATCCTCAAGAAATA 813
Db 2146 CAGGACCTGGCTGCTCCATACTTTCCTGTATAGTTCACGTTTCATGTAATCCTCAAGAAATA 2205
Qy 814 AAAGGAAGCCAAAGAG 829
Db 2206 AAAGGAAGCCAAAGAG 2221

RESULT 14
AX370851
LOCUS   AX370851   900 bp   DNA   linear   PAT 01-MAR-2002
DEFINITION Sequence 12 from Patent WO0210372.
ACCESSION   AX370851
VERSION   AX370851.1 GI:19168980
KEYWORDS   .
SOURCE   synthetic construct
  ORGANISM synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Chapman,P.W., de Luca,G. and Falciola,L.
  TITLE   Method of producing functional protein domains
  JOURNAL   Patent: WO 0210372-A 12 07-FEB-2002;
            Applied Research Systems ARS Holding S.A. (AN)
FEATURES   Location/Qualifiers
     source   1..900
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="mRNA for migSP fused to COL18A1 exons 38-41 for
                     pEnd-HR#1"

ORIGIN

Query Match   59.5%; Score 493.2; DB 6; Length 900;
Best Local Similarity   83.9%; Pred. No. 1.1e-55;
Matches 574; Conservative   0; Mismatches 98; Indels 12; Gaps   1;

Qy   1 CCCTGGCGGGCAGATGACATCCTGGCCGGCCGCCCCGGCCGCTGCTGGACCCCAGCCCTAC   60
Db 211 CCCTGGCGGGCAGATGACATCCTGGCCCAGCCCCCTGCCCGAGCCCCGAGCCCCAGCCCTAC 270
Qy   61 CCCGGGCCCCGCGACCACGGCTCCTACGGTGCACTTCCAGCCGGCTGCCCTCCCACTGGTGGG 120
Db 271 CCCGGAGCCCCGCCACCACCACGTCCTCCTACGGTGCACCTGCGCGGCGCGCGCGCCCACCAAG--- 326
Qy 121 CCCGTCCACACCCACCACACCCACCACCCCAGGACTTCCAGCTGCTGCGTGCACCTGGTGGCC 180
Db 327 ------CCCACCCGCCCACAGCCCACCGCGACTTCAGCCGCGTGCTTCCAGCCGGTTGCG 378
Qy 181 CTGAACAGCCCGAGCGCGGCGGCATGCGAGGCATCCGGGAGGCGGACTTCCAGTGCTTC 240
Db 379 CTCAACAGCCCCCTGTCAGCGGCGCATGCGGGGCATCCGGGGCATCCGCGGGGACTTCCAGTGCTTC 438
Qy 241 CAGCAGGCGCGCCGCGCGGGGCTGCCCGGCACCTTCCGGGCACCTTCCTGTCGTCGCCGGCTG 300
Db 439 CAGCAGGCGCGGCGGGCCGTGGGGCGGCCCGCACCTTCCCGCGGCCTTCCCGTCCTTCCGCGGCCTG 498
Qy 301 CAGGACCTTCTACAGCATCGTGCGCCGCTGCCGGGCGTCCCGGGGCACCGGGTGCCCGTCGTCAACCTC 360

```
                    /citation=[2]
                    /citation=[1]
                    /codon_start=1
                    /product="alpha 1(XVIII) collagen"
                    /protein_id="AAA20657.1"
                    /db_xref="GI:487734"
                    /translation="DSNAFVESGRPGLPGQGVVQGPSGPSPKGDKGEVGPPGPPGQPPID
                    LFHLRAEMKGDKGDRGDAGQKGERGEPGAPGGGFFSSSVPGPPGPGYPGIPGPKGES
                    IRGPPGPPGRQPPGIGYEGRQGPPGPGPPGPPGPGHRQTVSVPGPPGPPGP
                    PGAMGASAGQVRIWATYQTMLDKIREVPEGWLIFVAEREELYVRVNGPRKVLLEART
                    ALPRGTGNEVAALQPPLVQLHEGSPYTRREYSYSTARPWRADDILANPPRLPDRQPYP
                    GVPHHHSSYVHLPPARPTLSLAHTHQDFQPVLHIVALNTPLSGGMRGIRGADFQCFQQ
                    ARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIVNLKDEVLSPSWDSLFSGSGQL
                    QPGAR.TFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLMESYCETWRTETTGATGQASSL
                    LSGRLLEQKAASCHNSYIVLCIENSFMTSFSK"
     3'UTR          1450..1615
                    /citation=[2]
     polyA_signal   1584..1589
                    /citation=[2]
ORIGIN

Query Match       52.7%;   Score 436.8;   DB 10;   Length 1615;
Best Local Similarity   73.5%;   Pred. No. 2.4e-48;
Matches 625;   Conservative   0;   Mismatches 192;   Indels   33;   Gaps   4;

Qy      1  CCCTGGGCGGGCAGATGACATCCTGGCCCGGCCCCCCGCGCTGCTGACCCCCAGCCCTAC    60
           ||||||||| |||| |||||||||||||||||  ||||| || ||||||||||||||||
Db    766  CCCTGGGCGGAGCAGATGACATCCTGGCCAACCACCGGCGCCTGCCCAGACCCAGCCCTTAC   825

Qy     61  CCCGGGG---CCCGGCACCACGGCTCCTACGTGCACTTCAGCCGGCTCGCCCCACTGGT   117
           |||||||       ||||| |||||||||||||||  |||| |||||||||||||||
Db    826  CCCGGAGTTCCACATCACCACAGTTCCTATGTGCACCTGCCGCCCGCCCCAC----    881

Qy    118  GGGCCCGTCCACACCCACCCACCCACCCAGGACTTCCAGCTGGTGCGCCACCTGGTG   177
                 |||  |||  || |||||||| |||||||| |||| ||||| ||||||||||
Db    882  ------CCTCTCACTTGCTCATCATACTCATCAGGACTTTCAGCCAGTGCTCCCACCTGGTG   933

Qy    178  GCCCTGAACAGCCCGCAGCCGGCGGCCATGCGGCGAGGCATCCGGGGAGGCGAGACTTCCAGTGC   237
           || |||||  |||| || || |||| |||||  | ||||||||||  | ||||||||||||||||
Db    934  GCACTGAAACACCCCCCCTGTCTGAGGCATGCGTGGTATCCGTGGACAGCAGATTTCCAGTGC   993

Qy    238  TTCCAGCAGGCGGCGGCGCCGGCGGGCGTGGCCGGCGCACCTTCCGGGCCTTCCGTGTCGTGCGCGG   297
           |||||||| ||  |||| |||  || || ||| || ||||||||||||| |||||||||| || | ||
Db    994  TTCCAGCAAGCCCGAGCCCGGGACGTCGGCGAGCCTGTGTCGGCACCTTCCGGGCTTTCCGTGTCCTGTCCTCTAGG   1053

Qy    298  CTGCAGGACCTCTACAGCATCGTGCCGCCGCCGCCGCGACCGGGGTGCCCGTCGTCAAC   357
           ||||||||  ||||| |||||||| | |  || |||  ||  |||||  | || |||||
Db   1054  CTGCAGGATCTCTATAGCATCGTGCGCGCGTGCTGCTGCAGTGTCTGTGCCCATCGTCAAC   1113

Qy    358  CTCAGGGACGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTCCGGCTCCGAGGGCCAG   417
           || ||||||||||||||| | ||||||||||| | |  |||||| |||||| ||||| ||
Db   1114  CTGAAGGACGAGGTGCTATCTCCCAGCTGGGACTCCCTGTTTTCTGGCTCCCAGGGTCAA   1173

Qy    418  CTGAAGCCCCGGGGCCCGCATCTTCTCTTTCGACGGCCAGAGATGTCCTGCAGCACCCCGCC   477
           |||  ||||||||||||||||||||| |||||| | | |||||||||||||  ||| ||||
Db   1174  CTGCAACCCCGGGGCCCGCATCTTCTTTTTCGTTTGACGGCAGAGATGTCCTGAGACACCCCAGCC   1233

Qy    478  TGGCCCCGGAAGAGCGTGTGGCACGGCTCCGACAGCGGCGCGCCCCGGCCGGACACGAGC   537
           ||||||  |||||||| ||||||||||| ||| | ||| | |||||||  ||| | |  |
Db   1234  TGGCCCGAGAAGAGCGTATGGCACGGCTCGGACCTCGGACCCCAGTGGGCGGAGGCTGATGGGAGAGT   1293

Qy    538  TACTGCGAGACGTGGCGACGGAGGCCCCGGGGCGCCACCGGACGCGTCGTCGCTGCTG   597
           ||||| ||||| ||| || |||  |  |||||| |||||||  | |||||| |||||||
Db   1294  TACTGTGAGACATGCGAACTGGCAACTGAAACTACTGGGGCTACAGTCAGGCCTTCCTCGTCCTGCTG   1353

Qy    598  GCGGGCAGGCTGCTGGGAGCAGGGCCGGAGGCCGGAGGGCGCCACCGGCCTTCGGTGCTCTGC   657
           || ||||||||  ||||| |||  | ||| | ||| |   | | || |||| ||| |||| |
Db   1354  TCAGGCAGGCTCCCTGGAACAGAAACAGAAAGCTGCGACTTGCCACAACAGCTACATCGTCCTGTGTC   1413

Qy    658  ATCGAGAACAGCGTCATGACCTCCTTCTCCAAGTAGGCCGCGCCCACGGACAGGCG   717
           || ||||| ||| ||||||||| |||||||||| ||| || | || || |  |||
Db   1414  ATTGAGAATAGCTTCATGACCTCTTTCTCCAAATAGGGCCCTCTGCCAGCTAGGGTGGCAG   1473

Qy    718  GGGGAGG-----------------GGGCGGCCCGCAGGAGCATCCGCGCCCCGGGGGGGCC   761
           ||||                            ||| ||| |||  | ||||  |
Db   1474  ACAGAGGCCATGCAGAACTTGACACAGCGCAGGGAGCATTCAGTCAACCTCAGCACCCAGGGCTC   1533

Qy    762  TGGCCCGGACGCCTTGCCTCGCACCGT--CACGTTTAATGTAATCCTCAAGAAATAAAAGGA   819
           ||||  |||  |  |||| | ||| |    ||||| ||||| ||||||||||||| |||| |
Db   1534  TGGCTGGGATACAATCCTGTATAGTTCCCATTTTTATGTAATCCTCAAGAAATAAAAGGA   1593

Qy    820  AGCCAAAGAG   829
           ||||||||||
Db   1594  AGCCAAAGAG   1603

Search completed: March 29, 2004, 13:38:49
Job time : 3345.77 secs
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        March 29, 2004, 09:08:09 ; Search time 369.576 Seconds
                                 (without alignments)
                                 9529.183 Million cell updates/sec

Title:         US-09-938-391-1
Perfect score: 829
Sequence:      1 ccctggcgggcagatgacat.........aataaaggaggccaaagag 829

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:      6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    N_Geneseq_29Jan04:*
              1: geneseqn1980s:*
              2: geneseqn1990s:*
              3: geneseqn2000s:*
              4: geneseqn2001as:*
              5: geneseqn2001bs:*
              6: geneseqn2002s:*
              7: geneseqn2003as:*
              8: geneseqn2003bs:*
              9: geneseqn2003cs:*
             10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                            SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description		
1	829	100.0	829	6	AAL46062	Aal46062 Canine pr		ABA00774	Aba00774	Human end
2	555	66.9	555	6	AAL46063	Aal46063 Canine en		AAX35375	Aax35375	SEQ ID 50
3	550.4	66.4	552	3	AAZ51309	Aaz51309 Canine an		AAA68203	Aaa68203	Human end
4	503.2	60.7	4875	6	ABQ54955	Abq54955 Human ova		AAC62023	Aac62023	Nucleotid
5	503.2	60.7	5408	2	AAX77720	Aax77720 Human alp		AAC88289	Aac88289	cDNA enco
6	496.8	59.9	3394	2	AAT84484	Aat84484 Human alp		ABK50685	Abk50685	cDNA enco
7	496.8	59.9	3394	2	AAX78379	Aax78379 Human alp		AAH79104	Aah79104	Human end
8	496.8	59.9	3394	6	ABN95680	Abn95680 Gene #217		AAL55454	Aal55454	Specific
9	493.2	59.5	900	6	ABK09977	Abk09977 Synthetic		ABQ76079	Abq76079	Anticance
10	493.2	59.5	4551	6	ABN85301	Abn85301 Human col		AAL44000	Aal44000	Reproduct
11	493.2	59.5	4551	6	ABV94763	Abv94763 Human pan		ABQ81193	Abq81193	Human end
12	425.6	51.3	968	3	AAC62025	Aac62025 Nucleotid		AAA64014	Aaa64014	Nucleotid
13	423.6	51.1	1564	6	ABQ76740	Abq76740 DNA encod		ADE40255	Ade40255	Human NOV
14	423	51.0	4031	2	AAT84485	Aat84485 Mouse alp		AAA27005	Aaa27005	Alternate
15	421.2	50.8	816	3	AAA64013	Aaa64013 DNA encod		AAS00868	Aas00868	Human gen
16	420	50.7	555	4	AAI66529	Aai66529 Human vas		ABK09978	Abk09978	Synthetic
17	419.2	50.6	558	6	ABA99261	Aba99261 Human end		AAZ50398	Aaz50398	Human end
18	418.8	50.5	563	2	AAZ08750	Aaz08750 Human end		AAC67777	Aac67777	Murine en
19	418	50.4	546	3	AAA27004	Aaa27004 Human end		ABK47719	Abk47719	cDNA enco
20	418	50.4	549	3	AAX77719	Aax77719 Human end		AAX58740	Aax58740	DNA encod
21	418	50.4	549	3	AAZ51291	Aaz51291 Human ang		AAZ08748	Aaz08748	Mouse end
22	418	50.4	549	3	AAA29884	Aaa29884 Human ang		AAD18701	Aad18701	Mouse end
23	418	50.4	549	5	AAS00867	Aas00867 Human gen				
24	418	50.4	549	7	ABA00774					
25	418	50.4	552	2	AAX35375					
26	418	50.4	552	3	AAA68203					
27	418	50.4	552	3	AAC62023					
28	418	50.4	552	5	AAC88289					
29	418	50.4	552	6	ABK50685					
30	417.2	50.3	641	4	AAH79104					
31	417.2	50.3	641	6	AAL55454					
32	417.2	50.3	641	6	ABQ76079					
33	417.2	50.3	641	6	AAL44000					
34	417	50.3	551	5	ABQ81193					
35	412.2	49.7	6462	3	AAA64014					
36	411.6	49.7	555	9	ADB40255					
37	410.8	49.6	534	5	AAA27005					
38	410.8	49.6	537	6	AAS00868					
39	398.4	48.1	650	6	ABK09978					
40	394.2	47.6	552	3	AAZ50398					
41	376.8	45.5	573	3	AAC67777					
42	376.8	45.5	573	6	ABK47719					
43	376.6	45.4	565	2	AAX58740					
44	376.4	45.4	580	2	AAZ08748					
45	376.2	45.4	558	4	AAD18701					

```
                              ALIGNMENTS

RESULT 1
AAL46062
ID   AAL46062 standard; cDNA; 829 BP.
XX
AC   AAL46062;
XX
DT   19-JUL-2002 (first entry)
XX
DE   Canine pro-endostatin coding sequence.
XX
KW   Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;
KW   psoriasis; rheumatoid arthritis; retinopathy; macular degeneration;
KW   corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
KW   rubeosis; Osler-Webber Syndrome; myocardial angiogenesis;
KW   plaque neovascularisation; telangiectasia; haemophiliac joints;
KW   angiofibroma; wound granulation; coronary collateral;
KW   cerebral collateral; arteriovenous malformation;
KW   ischaemic limb angiogenesis; diabetic neovascularisation; fracture;
KW   cytostatic; antiarthritic; antipsoriatic; antidiabetic; ophthalmological;
KW   gynaecological; gene; ss.
XX
OS   Canis familiaris.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..693
FT                   /tag= a
FT                   /product= "pro-endostatin"
FT                   /partial
FT                   /note="no start codon"
XX
PN   EP191036-A2.
XX
PD   27-MAR-2002.
XX
PF   24-AUG-2001; 2001EP-00307224.
XX
PR   25-AUG-2000; 2000US-0227924P.
XX
PA   (PFIZ ) PFIZER PROD INC.
XX
PI   Sheppard MG, Tong X;
XX
DR   WPI; 2002-354068/39.
DR   P-PSDB; AA017429.
XX
PT   An isolated nucleic acid molecule for the treatment of angiogenesis-
```

PT related disorder, such as cancers or diabetic retinopathy, encodes an
PT endostatin protein.
XX
PS Claim 2; Fig 2; 56pp; English.
XX
CC The present invention provides the protein and coding sequences of canine
CC pro-endostatin and endostatin. The sequences can be used in the treatment
CC and diagnosis of angiogenesis related disorders, including cancer,
CC rheumatoid arthritis, psoriasis, retinopathy, macular degeneration,
CC corneal graft rejection, neovascular glaucoma, retrolental fibroplasia,
CC rubeosis, Osler-Webber Syndrome, myocardial angiogenesis, plaque
CC neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,
CC wound granulation, coronary collaterals, cerebral collaterals,
CC arteriovenous malformations, ischaemic limb angiogenesis, diabetic
CC neovascularisation, and fractures. The present sequence is the canine pro
CC -endostatin coding sequence
XX
SQ Sequence 829 BP; 124 A; 314 C; 278 G; 113 T; 0 U; 0 Other;

Query Match   100.0%;   Score 829; DB 6; Length 829;
Best Local Similarity 100.0%;   Pred. No. 1.7e-126;
Matches 829; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy     1 CCCTGGGGGGCAGATGACATCCTGGCCGGCCCCCGCCGCCTGGACCCCCAGCCCTAC  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CCCTGGGGGGCAGATGACATCCTGGCCGGCCCCCGCCGCCTGGACCCCCAGCCCTAC  60

Qy    61 CCCGGGGCCCCGCACCACGGCTCCTACGTGCACTTCCAGCCGGCTGGCCCCACTGGTGGG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CCCGGGGCCCCGCACCACGGCTCCTACGTGCACTTCCAGCCGGCTCGCCCCACTGGTGGG 120

Qy   121 CCCGTCCACACCCACCACCACCAGGACTTCAGCTGGTGCTGCACCTGGTGGCC 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CCCTCCACACCCACCACCACCAGGACTTCAGCTGGTGCTGCACCTGGTGGCC 180

Qy   181 CTGAACAGCCCGGCAGCCGGCCGGCATCGGAGGCATCCGGGGAGCCGGACTTCCAGTGCTTC 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGAACAGCCCGGCAGCCGGCCGGCATCGGAGGCATCCGGGGAGCCGGACTTCCAGTGCTTC 240

Qy   241 CAGCAGGCGCGCGCCGCGGGGGCTGGCCGCCACCTTCGGGCCTTCGTCGTCGGCGTG 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CAGCAGGGGCGCGCCGGGGGGCTGGCCGCGGCACCTTCGGGCCTTCGTCGTCGGCGTG 300

Qy   301 CAGGACTTCTACAGCATCGTGCGCCGGCATCGTGCGCCGGGGTGCCCGTCGTCAACCTC 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CAGGACCTTCTACAGCATCGTGCGCCGGCATCGTGCGCCGGGGTGCCCGTCGTCAACCTC 360

Qy   361 AGGGACGAGGTGCTCTCCCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTG 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AGGGACGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTG 420

Qy   421 AAGCCCGGGGCCCGCATCTTCTTTTGACGGCAGAGATGTCCTTGCACCACCCCGCCTGG 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 AAGCCCGGGGCCCGCATCTTCTTTTGACGGCAGAGATGTCCTTGCAGCACCCCGCCTGG 480

Qy   481 CCCCGGAAGAGCGTGTGGCACGGCTCCGAACCCCAGCGGGCGCCGCCTGACCAGCTAC 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CCCCGGAAGAGCGTGTGGCACGGCTCCGAACCCCAGCGGGCGCCGCCTGACCAGCTAC 540

Qy   541 TGCCGAGACGTGGGGGACGGAGGGCCCCGGCCAGGCGGTCGTCGCTGCTCGCG 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TGCCGAGACGTGGGGGACCGGGGACGGAGGGCCCCGGCCAGGCGGTCGTCGCTGCTCGCG 600

Qy   601 GGCAGGTGCTGAGCAGCAGGAGGCCGCGAGTGCCGCCACGCCTTCGTGGTGCTCTGCATC 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GGCAGGTGCTGAGCAGCAGGAGGCCGCGAGCTGCCGCCACGCCTTCGTGGTGCTCTGCATC 660

Qy   661 GAGAACAGCGTCATGACCTCCTTCTCCAAGTAGGGCCCGCCACCGCACAGGCGGGG 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 GAGAACAGCGTCATGACCTCCTTCTCCAAGTAGGGCCCGCCACGGACAGGCGGGG 720

Qy   721 GAGGGGCGCCGCCCCGCAGGAGCATCCGCCGCCCCGGGGGGGGGGGGCCTGGCCGGACGCTTGCCTG 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GAGGGGCGCCGCCCCGCAGGAGCATCCGCCGCCCCGGGGGGGGGGGGCCTGGCCGGACGCTTGCCTG 780

Qy   781 CACCGTCACGTTAATGTAATCCTCAAGAAATAAAAGGAAGCCAAAGAG 829
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   781 CACCGTCACGTTAATGTAATCCTCAAGAAATAAAAGGAAGCCAAAGAG 829
```

RESULT 2
AAL46063
ID   AAL46063 standard; cDNA; 555 BP.
XX
AC   AAL46063;
XX
DT   19-JUL-2002   (first entry)
XX
DE   Canine endostatin coding sequence.
XX
KW   Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;
KW   psoriasis; rheumatoid arthritis; retinopathy; macular degeneration;
KW   corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
KW   rubeosis; Osler-Webber Syndrome; myocardial angiogenesis;
KW   plaque neovascularisation; telangiectasia; haemophiliac joints;
KW   angiofibroma; wound granulation; coronary collateral;
KW   cerebral collateral; arteriovenous malformation;
KW   ischaemic limb angiogenesis; diabetic neovascularisation; fracture;
KW   cytostatic; antiarthritic; antipsoriatic; antidiabetic; ophthalmological;
KW   gynaecological; gene; ss.
XX
OS   Canis familiaris.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..555
FT   /*tag= a
FT   /product= "endostatin"
FT   /partial
FT   /note= "no start codon"
XX
PN   EP1191036-A2.
XX
PD   27-MAR-2002.
XX
PF   24-AUG-2001; 2001EP-00307224.
XX
PR   25-AUG-2000; 2000US-0227924P.
XX
PA   (PFIZ) PFIZER PROD INC.
XX
PI   Sheppard MG, Tong X;
XX
DR   WPI; 2002-354068/39.
DR   P-PSDB; AAO17430.
XX
PT   An isolated nucleic acid molecule for the treatment of angiogenesis-
PT   related disorder, such as cancers or diabetic retinopathy, encodes an
PT   endostatin protein.
XX
PS   Claim 2; Fig 4; 56pp; English.
XX
CC   The present invention provides the protein and coding sequences of canine
CC   pro-endostatin and endostatin. The sequences can be used in the treatment
CC   and diagnosis of angiogenesis related disorders, including cancer,
CC   rheumatoid arthritis, psoriasis, retinopathy, macular degeneration,
CC   corneal graft rejection, neovascular glaucoma, retrolental fibroplasia,
CC   rubeosis, Osler-Webber Syndrome, myocardial angiogenesis, plaque
CC   neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,
CC   wound granulation, coronary collaterals, cerebral collaterals,
CC   arteriovenous malformations, ischaemic limb angiogenesis, diabetic
CC   neovascularisation, and fractures. The present sequence is the canine
CC   endostatin coding sequence
XX
SQ   Sequence 555 BP; 78 A; 203 C; 191 G; 83 T; 0 U; 0 Other;

Query Match   66.9%;   Score 555;   DB 6;   Length 555;

Best Local Similarity   100.0%; Pred. No. 7.8e-82;
Matches 555; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   139 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 198
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG   60

Qy   199 GGCGGCATGCGGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGGCGCGCCGCCG 258
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGCGGCATGCGGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGGCGCGCCGCCG 120

Qy   259 GGGCTGGCCGGCCACCTTCCGGGCCTTCCTGTCGCGGCTGCGAGGACCTCTACAGCATC 318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGGCTGGCCGGCCACCTTCCGGGCCTTCCTGTCGCGGCTGCGAGGACCTCTACAGCATC 180

Qy   319 GTGCGGCCGCGCCGCCACCGGCGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTGCGGCCGCGCCGCCACCGGCGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 240

Qy   379 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCCGATC 438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCCGATC 300

Qy   439 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGGCCCCGGAAGAGCGTGTGG 498
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGGCCCCGGAAGAGCGTGTGG 360

Qy   499 CACGGCTCCGACCCCAGCCGGCGGCGCCCCGACCGACAGCTACTGCGAGACGTGGCCGACG 558
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CACGGCTCCGACCCCAGCCGGCGGCGCCCCGACCGACAGCTACTGCGAGACGTGGCGACG 420

Qy   559 GAGGCCCCGGCGGCCACCGGGCCAGGCTCGTCGCTCGCTGCGGGGCAGGCTGCTGGAGCAG 618
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAGGCCCCGGCGGCCACCGGGCCAGGCGTCGTCGCTCGCTGCGGGGCAGGCTGCTGGAGCAG 480

Qy   619 GAGGCCGCGAGCTGCCGCCACGCCTTCGTTGGTGCTCGATCGAGAACACAGGCTCATGACC 678
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAGGCCGCGAGCTGCCGCCACGCCTTCGTTGGTGCTCGTGTGCTCATCCGAGAACAGCCTCATGACC 540

Qy   679 TCCTTCTCCAAGTAG 693
           |||||||||||||||
Db   541 TCCTTCTCCAAGTAG 555

RESULT 3
AAZ51309
ID   AAZ51309 standard; DNA; 552 BP.
XX
AC   AAZ51309;
XX
DT   06-JUN-2000 (first entry)
XX
DE   Canine angiogenesis inhibitor, endostatin DNA.
XX
KW   Canine; immunoglobulin Fc fragment; endostatin; immunofusin;
KW   angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;
KW   antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;
KW   vasotropic; vulnerary; treatment; antiarteriosclerosis; tumour;
KW   metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
KW   ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW   myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW   wound granulation; keloid scar; gene therapy; ds.
XX
OS   Canis familiaris.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..552
FT   /tag= a
FT   /product= "Endostatin"
FT   /note= "Does not include stop codon."
FT   /partial
XX
PN   WO200011033-A2.

XX
PD   02-MAR-2000.
XX
PF   25-AUG-1999;   99WO-US019329.
XX
PR   25-AUG-1998;   98US-0097883P.
XX
PA   (LEXI-) LEXINGEN PHARM CORP.
XX
PI   Lo K, Li Y, Gillies SD;
XX
DR   WPI; 2000-237616/20.
DR   P-PSDB; AAY70265.
XX
PT   Novel fusion protein of angiostatin or endostatin and an immunoglobulin
PT   FC region, useful for treating conditions mediated by angiogenesis, such
PT   as rheumatoid arthritis, tumors and macular degeneration.
XX
PS   Example 8; Page 58-59; 68pp; English.
XX
CC   The patent discloses a DNA molecule encoding a fusion protein comprising
CC   a signal sequence, an immunoglobulin Fc region, and an angiogenesis
CC   inhibitor selected from angiostatin, endostatin, a plasminogen fragment
CC   having angiostatin activity, a collagen XVIII fragment having endostatin
CC   activity, or combinations of them. The fusion protein (immunofusin) is
CC   used to inhibit angiogenesis and to treat diseases or conditions mediated
CC   by angiogenesis. Conditions that may be treated include solid tumours,
CC   blood born tumours, tumour metastasis, benign tumours including
CC   haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic
CC   granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
CC   e.g. diabetic retinopathy, retinopathy of prematurity, macular
CC   degeneration, corneal graft rejection, neovascular glaucoma, retrolental
CC   fibroplasia, rubeosis and Osler-Webber syndrome; myocardial angiogenesis,
CC   plaque neovascularisation, telangiectasia, haemophiliac joints;
CC   angiofibroma, wound granulation, and excessive or abnormal stimulation of
CC   endothelial cells, intestinal cells, atherosclerosis, scleroderma and
CC   hypertrophic scars, i.e. keloid scars. The DNA constructs may be used in
CC   gene therapy. The present sequence is a DNA encoding canine endostatin
CC   used in the construction of immunofusin containing canine immunoglobulin
CC   Fc fragment
XX
SQ   Sequence 552 BP; 77 A; 204 C; 190 G; 81 T; 0 U; 0 Other;

Query Match   66.4%; Score 550.4; DB 3; Length 552;
Best Local Similarity   99.8%; Pred. No. 4.4e-81;
Matches 551; Conservative   0; Mismatches   0; Indels   1; Gaps   0;

Qy   139 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 198
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG   60

Qy   199 GGCGGCATGCGGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGGCGCGCCGCCG 258
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGCGGCATGCGGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGGCGCGCCGCCG 120

Qy   259 GGGCTGGCCGGCCACCTTCCGGGCCTTCCTGTCGCGGCTGCGTGCGGCTGCGAGGACCTCTACAGCATC 318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGGCTGGCCGGCCACCTTCCGGGCCTTCCTGTCGCGGCTGCGTGCGGCTGCGAGGACCTCTACAGCATC 180

Qy   319 GTGCGGCCGCGCCGCCACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTGCGGCCGCGCCGCCACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 240

Qy   379 CCCAGCTGGGAGGCCTTATTCCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC 438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CCCAGCTGGGAGGCCTTATTCCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC 300

Qy   439 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGGCCCCGGAAGAGCGTGTGG 498
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGGCCCCGGAAGAGCGTGTGG 360

Qy   499 CACGGCTCCGACCCCAGCCGGCGGCGCCCTGACCGACAGCTACTGCGAGACGTGGCGGACG 558
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db   361 CACGGCGTCCGACCCCAGCGGCGGCCGCCGCCGCCTGACCGACAGCTACTGCGAGACGTGGCGGACG 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   559 GAGGCCCCGGCGGGCCACCGGGCGAGGCGTCGTCGCTGCTGGCCGCGCAGGCCTCCTGGAGCAG 618
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAGGCCCCGGCGCAGGCGTCGTCGCTGCTGGCGGGCAGGCCTGCTGGAGCAG 480
             |||||||||||||||||||||||||||||||||||||||||||
Qy   619 GAGGCCGGAGCGTCCGCCACCGCCTTCGTCGGTCGCTCTGCCATCGAGAACAGCGGTCATGACC 678
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAGGCCGCCGAGCTGCCGCCACCGCCTTCGTCGGTGTCGCTCTGCATCGAGAAACAGCGTCATGACC 540
             |||||||||||||||||||||||||
Qy   679 TCCTTCTCCAAG 690
             |||||||||||
Db   541 TCCTTCTCCAAG 552
```

RESULT 4
ABQ54955
ID   ABQ54955 standard; cDNA; 4875 BP.
XX
AC   ABQ54955;
XX
DT   22-AUG-2002 (first entry)
XX
DE   Human ovarian antigen HEEBK29 cDNA, SEQ ID NO:835.
XX
KW   Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW   ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW   infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW   PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW   inflammatory condition; immune disorder; blood disorder;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disorder; urinary system disorder; drug screening;
KW   gene therapy; chromosome mapping; forensic analysis;
KW   antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW   antiinflammatory; gynaecological; reproductive; chromosome 21q22.3; gene;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO20020677-A1.
XX
PD   03-JAN-2002.
XX
PF   07-JUN-2001; 2001WO-US018569.
XX
PR   07-JUN-2000; 2000US-0209467P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE , Rosen CA;
XX
DR   WPI; 2002-147878/19.
DR   P-PSDB; ABP41878.
XX
PT   Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT   useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
PT   cancer), immune disorders, cardiovascular disorders and neurological
PT   diseases.
XX
PS   Claim 1; SEQ ID NO 835; 2922pp; English.
XX
CC   The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC   ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC   encompasses polypeptides 90% identical and polynucleotides 95% identical
CC   to the sequences of the invention. The invention additionally relates to
CC   recombinant vectors and host cells comprising human ovarian antigen
CC   polynucleotides, antibodies against human ovarian antigens, and the use
CC   of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC   treating, prognosing or preventing various ovary and/or breast-related
CC   disorders. Such conditions include ovarian cancer and breast cancer, and
CC   metastatic tumours of ovarian or breast origin, reproductive system
CC   disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC   polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC   disorders, infections (e.g. chlamydia, HIV, toxoplasmosis, and toxic
CC   shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC   vaginitis), immune disorders (e.g., congenital and acquired
CC   immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC   blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC   respiratory disorders, neurological disorders, gastrointestinal disorders
CC   and urinary system disorders. Ovarian antigen polypeptides and
CC   polynucleotides may also be used in screening for compounds which
CC   modulate ovarian antigen expression or activity. The polynucleotides may
CC   further be used for gene therapy, chromosome mapping, in the
CC   identification of individuals and in forensic analysis, and the
CC   polypeptides may be used as food additives or to prepare antibodies
CC   useful in disease diagnosis, drug targeting and phenotyping. The present
CC   sequence represents cDNA encoding a human ovarian antigen of the
CC   invention. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4875 BP; 892 A; 1569 C; 1570 G; 829 T; 0 U; 15 Other;

Query Match   60.7%; Score 503.2; DB 6; Length 4875;
Best Local Similarity   77.6%; Pred. No. 1.8e-73;
Matches 664; Conservative   0; Mismatches 153; Indels 39; Gaps   3;

```
Qy      1 CCCTGCGGGCAGATGACATCCTGCCCGCCCCCGCGCCTGCTGGACCCCAGCCCTAC 60
            ||||||||||||| ||||||||| ||||||||||| |||||||| |||||||||
Db   2386 CCCTGCGGGCAGATGACATCCTGCCCAGCCCCCTGCCCGCCCCAGCCCTAC 2445

Qy     61 CCCGGGCCCCGCACCAGGCTCCTACGTGCACTTCAGCGCGGCTCGCCCCACTGTGGG 120
            ||||||||| |||||||||| |||||| ||||||||||||||| ||||
Db   2446 CCCGGAGCCCCGCACCACAGTCCTACGTGCACGTGCACCTGGCGCRCCAGCCCACCAAG---- 2501

Qy    121 CCGTCCACACCCACACCCACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCC 180
            ||||||||||||||||| |||||| |||||||||||||||||||||||||||
Db   2502 -------CCCACCCGCCCACAGCCACCCGGACTTCCAGCCGGTGCTGCACCTGGTTGCG 2553

Qy    181 CTGAACAGCCCGCAGCCCGGCGGCGGCCATGCGGAGGCATCGGGAGGCGGACTTCCAGTGCTTC 240
            ||||||||||||| ||||| ||||||||||| |||||||||||||| |||||||||||
Db   2554 CTCAACAGCCCCCTGTCAGGCGGCATGCGGCCGCATCCGCGGGGGCCGACTTCCAGTGCTTC 2613

Qy    241 CAGCAGGCGGCGCGCCCGGCGGGCTGGCCGCGGCCACCTTCCGGGCTTCGGCGTCGGTGTCGGCGG 300
            ||||||||||||| ||||| ||||||||| ||||||| ||| |||||||| |||||
Db   2614 CAGCAGGGCGCGGCGCGCCGTGGGGCTGGCGGCGGCCACCTTCCGGGCTTCCTCCGCGGCCTG 2673

Qy    301 CAGGACCTCTACAGCATCGTGCCCGCCGCGGCACCGGGGTGCCGTCGTCAACCTC 360
            ||||||||||||| |||||| |||||| ||||||||||||| ||||||||||
Db   2674 CAGGACCTGTACAGCATCGTGCCGCCGTGCCGACCGCCAGCGTGCCCATCGTCAACCTC 2733

Qy    361 AGGGACGAGGTGCTCTTCCCCCAGCTGGGAGGCCCTTATTCTCGGGAGGTCCAGAGGGCCAGCTG 420
            |||||||||||||| ||||||| ||||||| |||||| || ||||||||||||||||||||
Db   2734 AAGGACGAGCTGCTGTTTTCCCAGCTGGAGGCTCTGTTCTTCAGGCTCTGAGGGTCCGCTG 2793

Qy    421 AAGCCCGGGCCCCCATCTCTCTTCCACGGCGAGAGATGTCTGCAGCGCCCGGCCTGG 480
            ||||||||| |||| ||||||||| |||||||| |||||||| ||||||||||||||
Db   2794 AAGCCCGGGCCAGCGCACGCATCTTCTCCCTTTGACGGCAAGGACGTCCTGAGGCACCCCACCTGG 2853

Qy    481 CCCCGGAAGAGAGCGTGTGGCACGGCTCCGACCCCAGCCGGGGGCGCCTGACCAGACAGCTAC 540
            |||||||||||| ||||||||||| |||||||||| ||||||||||||| ||| |||||
Db   2854 CCCCAGAAGAGAGCGTGTGGCATGGCTCGGACCCCCAACGGACCCCCAGAGGGCCTGACCGAGAGCTAC 2913

Qy    541 TGCGAGACGTGGCGGACGAGACGGCCCGCGGGCCACCGGGCAGGCGGTCGTGCTGCTGGGCG 600
            ||||||| ||||||||||| |||||||||| |||||||| ||||||||||||||||||||
Db   2914 TGTGAGACGTGGCGGACCGGAGGCGTCCCTCGGCCACCGGGCAGGCCTTCCTGCTGCTGGGGG 2973

Qy    601 GGCAGGCTGCTGAGCAGGAGGCCGGAGCTGCGGCCACCGGCCTTCGTGTTGGTCTTGCATC 660
            |||||| ||||||||||||||||||||||||| |||||||||| |||||||| |||||
Db   2974 GGCAGGCTCCTGGGGCAGCAGTGCGGCGAGCTGCCATCACGCCTACATCGTCTCTGCATT 3033

Qy    661 GAGAACAGCGTCATGACCTGCTTCTCCAAGTAGGGCCGGCGGGCCCACGGACAGCAGGCGGG 720
            ||||| ||| ||||| ||||||||| ||||||| |||||||||||||| |||||
Db   3034 GAGAAACAGCTTCATGACTGCCTCCAAGTAGCTGCCCCGCCTGGATGCRGARGTCCCGAGAGG 3093

Qy    721 GAGGGGGGGCCCCGGCCCCAGGAGCATCCGCGGCCCCGGGG-------------------- 756
```

Mon Mar 29 17:32:18 2004

us-09-938-391-1.rng

Page 5

```
Db   3094  ACCGGCGGCTCGGAGGAAGCCCCACCGTGGGCAGGGAGCRGCCGGCGGCCCTGGCCC  3153
Qy    757  --GGGGCTGGCCGGGACGCGCTTGCCTGCCACCG-TCACGTTTAATGTAATCCTCAAGAAATA  813
Db   3154  CAGGACCTGGCTGCCATACTTTCCTGTATAGTTCACGTTTCATGTAATCCTCAAGAAATA  3213
Qy    814  AAAGGAAGCCAAAGAG  829
Db   3214  AAAGGAAGCCAAAGAG  3229
```

RESULT 5
AAX77720
ID   AAX77720 standard; DNA; 5408 BP.
XX
AC   AAX77720;
XX
DT   10-AUG-1999 (first entry)
XX
DE   Human collagen 18 DNA.
XX
KW   Plasminogen; human; angiostatin; endostatin; gene therapy; collagen 18;
KW   anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW   tumour growth; solid tumour; diabetic retinopathy; retina; vector; ss.
XX
OS   Homo sapiens.
XX
PN   WO9926480-A1.
PD   03-JUN-1999.
XX
PF   20-NOV-1998; 98WO-US024950.
XX
PR   20-NOV-1997; 97US-0097 5424.
XX
PA   (GENE-) GENETIX PHARM INC.
PA   (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
PI   Lebouleh P, Pawliuk RJ, Bachelot T;
XX
DR   WPI; 1999-357696/30.
DR   P-PSDB; AAY08694.
XX
PT   Anti-angiogenic gene therapy vectors.
XX
PS   Disclosure; Page 75-76; 83pp; English.
XX
CC   This invention describes a novel viral gene therapy vector comprising a
CC   nucleic acid molecule encoding an anti-angiogenic polypeptide chosen from
CC   human or murine angiostatin, human or murine endostatin and angiogenesis-
CC   inhibiting fusions and fragments, where the viral vector is sufficiently
CC   attenuated for use in human gene therapy. The products of the invention
CC   have anti-angiogenic, cytostatic, anti-diabetic and ophthalmological
CC   activity. The vector is used in gene therapy for inhibiting tumour growth
CC   in humans harbouring a solid tumour. The vector expresses an anti-
CC   angiogenic polypeptide. An additional use comprises treatment of diabetic
CC   retinopathy, where the anti-angiogenic polypeptide inhibits angiogenesis
CC   in the vicinity of the retina. The vector is administered to cells ex
CC   vivo and then administered to the patient
XX
SQ   Sequence 5408 BP; 986 A; 1820 C; 1796 G; 806 T; 0 U; 0 Other;

```
Query Match  60.7%;  Score 503.2;  DB 2;  Length 5408;
Best Local Similarity  77.6%;  Pred. No. 1.8e-73;
Matches 664;  Conservative 0;  Mismatches 153;  Indels 39;  Gaps 3;

Qy      1  CCCTGGCGGGCAGATGACATCCTGGCCCGCCCCCCGCCCTGCTGGACCCCCAGCCCTAC  60
Db   3393  CCCTGGCGGGCAGATGACATCCTGGCCCGCCCCCTGCCCTGCCCGAGCCCCAGCCCTAC  3452
Qy     61  CCCGGGGGCCCCGCACCACCGCTCACTTCCAGCCGCGGCTGCCCCCACTGGTGGG  120
```

```
Db   3453  CCCGGAGCCCCGCACCACAGTCTTCACGTCGCCGGCGGCGGCGGACCCACAAG----  3508
Qy    121  CCCGTCCACACCACCACCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCC  180
Db   3509  ------CCCACCCGCTCACCACCAGCGCCACCGGACTTCCAGCCGGTGCTCCACCTGGTTGCG  3560
Qy    181  CTGAACAGCCCGCCAGCGCGGCGCATCGGAGGCATCCGGGGGGGCGACTTCCAGTGCTTC  240
Db   3561  CTCAACAGCCCCCTGTCAGGCGGCATCGGGGGCCATCCGGGGGCCAGCTTCCAGTGCTTC  3620
Qy    241  CAGCAGGCGCGCGGCCGCGGGGGCTGGCCTGGCGGCGGCACCTTCCGGCGCCACCTG  300
Db   3621  CAGCAGGCGGCGGGGCCGTGGGGGTCGGGGCGGCACCTTCCGGCGCCATCGTCAACCTG  3680
Qy    301  CAGGACCTCTACAGCATCGTGCCGCGCGGGGTGCCCGTCGTCAACCTC  360
Db   3681  CAGGACCTGTACAGCATCGTGCCGCGCGGGGTGCCCATCGTCAACCTC  3740
Qy    361  AGGGACGAGGTGCTCTTCCCCAGCTGGGAGGCCTATTCTCGGGTCCAGGGGCCAGCTG  420
Db   3741  AAGGACGAGAGCTGCTCGTGTTCCCAGCTGGGAGGTCTGTTCTCAGGCTCTGAGGGTCCGGTG  3800
Qy    421  AAGCCCGGGCCGCCGCCATCTTCCTTTCGAGGCAGAGATGTTCCTGCGACACCCCGCCTGG  480
Db   3801  AAGCCCGGGGCCGCCATCTCCTTTGACGGCAAGGACGTTCCTGAGGGCACCCCACCTGG  3860
Qy    481  CCCCGGAAGACGTGTGGCACGGTCCGACCCCAGCGGCGGCGGCCGCCTGACCGACAGCTAC  540
Db   3861  CCCCAGAAGAGAGCGTGTGGCATGGCTCGGACCCCAACCGGGGCCAGGCGGCGGCAGCTAC  3920
Qy    541  TGCGAGACGTGCGACGGAGAGCCCCGGCGGCCGGCAGGCCGTCGTCGCTGCTGGCCG  600
Db   3921  TGTGAGACGTGGCGGACGAGAGGCTCCCGGCCAGGAGGCTCCTGCTGCTGGGG  3980
Qy    601  GGCAGGCGTGCTGGAGCAGGAGCAGGAGGGCCGGCGAGCTGCCGCCACCGCTTCGTGGTGCTCTGCATC  660
Db   3981  GGCAGGCGTCCTGGGGCAGAGTGCCGGGAGCTGCCATCACGCTACACATCGTGCTGCTGCATT  4040
Qy    661  GAGAACAGCGTCATGACCTCCTTCTTCCCAAGTAGGCCCGGCGCCCACGACAGCGCGGG  720
Db   4041  GAGAACAGCTTCATGACTGCCTCCAAGTAGCCACCGCCCTGATGCGGATGCCGGAGAGG  4100
Qy    721  GAGGGGGCGCCCCGCCAGGAGACCATCCGCCCGCCCGGGG--------  756
Db   4101  ACCGGCGGCTCGGAGGAAGCCCCACCGTGGGCAGGGAGGCGCCGGCCCCTGGCCC  4160
Qy    757  --GGGCCTGGCCGGGACGCGCTTGCCTGCCACCG-TCACGTTTAATGTAATCCTCAAGAAATA  813
Db   4161  CAGGACCTGGCTGCCATACTTTCCTGTATAGTTCACGTTTCATGTAATCCTCAAGAAATA  4220
Qy    814  AAAGGAAGCCAAAGAG  829
Db   4221  AAAGGAAGCCAAAGAG  4236
```

RESULT 6
AAT84484
ID   AAT84484 standard; cDNA; 3394 BP.
XX
AC   AAT84484;
XX
DT   19-NOV-1997 (first entry)
XX
DE   Human alpha-1 collagen cDNA.
XX
KW   Alpha-1 collagen; type XVIII collagen; cartilage degeneration; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..2055
FT   /*tag= a
XX

PN   US5643783-A.
XX
PD   01-JUL-1997.
XX
PF   01-DEC-1993;   93US-00159784.
XX
PR   01-DEC-1993;   93US-00159784.
XX
PA   (HARD) HARVARD COLLEGE.
XX
PI   Olsen BR, Oh SP;
XX
DR   WPI; 1997-350247/32.
DR   P-PSDB; AAW26327.
XX
PT   Nucleic acid encoding human alpha-1 collagen - for production of
PT   recombinant alpha-1 collagen, for use in the treatment of cartilage
PT   degeneration.
XX
PS   Claim 1; Col 23-30; 35pp; English.
XX
CC   A cDNA clone (AAT84484) codes for a human novel type alpha-1 (XVIII)
CC   collagen (AAW26327) that is expressed in multiple tissues, especially
CC   liver, lung and kidney. It was isolated from a placental cDNA library
CC   using a probe based on an unidentified collagenous protein and a probe
CC   based on mouse alpha-1 cDNA clone mc19. A claimed plasmid comprising
CC   alpha-1 collagen nucleic acid and an expression control sequence can be
CC   used to express recombinant collagen in prokaryotic or eukaryotic
CC   (especially mammalian) host cells. The collagen may be used to treat a
CC   patient suffering from a disease associated with degradation of
CC   cartilage, and for supplementing collagen
XX
SQ   Sequence 3394 BP; 668 A; 1130 C; 1080 G; 516 T; 0 U; 0 Other;

Query Match   59.9%; Score 496.8; DB 2; Length 3394;
Best Local Similarity   77.1%; Pred. No. 2.1e-72;
Matches 660; Conservative   0; Mismatches 157; Indels 39; Gaps   3;

Qy   1 CCCTGGCGGGCAGAGATGACATCCTGGCCCGCCCCGCTCTGCCTGACCCCCAGCCCTAC 60
          |||||||||||||| |||||||||||||||||| ||||| |||||||||||||||||
Db   1378 CCCTGGCGGGCAGATGACATCCTGGCCCGCCCCGGTGGCCCGGAGCCCCAGCCCTAC 1437

Qy   61 CCCGGGGCCCCGCACCACGGCTCCTACGGTGCACTTCCAGCCGGCTGCCCCCACTGGTGGG 120
          ||||| |||||||||| ||||||||||| || |||||||| |||| || |||||||
Db   1438 CCCGGAGGCCCGCACCACAGCTCCTACGTGCACTGCGGCCCCGGCACGACCCACCACAAG---- 1493

Qy   121 CCGTCCACACCCACCACACCCACCAGGACTTCCAGCTTGGTGCTGCACCTGGTGGCC 180
          | |||||| || | || ||| ||||||||||| || || ||||||||||
Db   1494 ---------CCACCCGCCCACAGCCACCGCACCGACTTCCAGCCGACTGCTGTCCACCTGGTTGCG 1545

Qy   181 CTGACACAGCCCGCAGCCGGGCGGCCGCGCATCCGGAGGCATCCGGGAGGCCGGACTTCCAGTGCTTC 240
          || | |||||| | ||||||||||| |||||||| |||||||| ||||||||||||||||||||||
Db   1546 CTCAACAGCCCCCTGTCAGGCGGCCATGGGGGCATCCGGGGGGCCGGACTTCCAGTGCTTC 1605

Qy   241 CAGCAGGCGGCGGCCGGCCGGGGCGTGGCCGGCACCTTCGTGCGTCGGGGCTG 300
          ||||||||||||||||| |||||||| ||||||||||||| ||| ||||||||
Db   1606 CAGCAGGCGGCGGCCGGGGCGTGGGGCTGGCGGGCACCTTCCGTGTCCTCGGCGGCCTG 1665

Qy   301 CAGGACCTTCTACAGCATCGTGCGCCGGGCCGGGCACCGGGGTGCCGTCGTCAACCTC 360
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1666 CAGGACCTTGTACAGCATCGTGCGCCGGGCCGGGCACCGGGGTGCCGTCGTCAACCTC 1725

Qy   361 AGGGACGAGGTGCTTCCCAGCTGGGAGGGCTTATTCTCGGGGCTTCCGAGGGCCAGCTG 420
          |||| ||||||||| | |||| ||||||| || || ||||||| ||||||||||||
Db   1726 AAGGACGAGCTGCTGCTGTTTCCCAGCTGGGAGGCTGGGCTCTGTTCTCAGGGGTCCGGCTG 1785

Qy   421 AAGCCCGGGCCCCGATCTTCTCTTCGACGGGAGAGATGTCCTGCAGCACCCCGCCTGG 480
          |||| ||||| | |||||| | | |||||| |||||||||||||||||||| |||
Db   1786 AAGCCCCGGGCACGCATCTTCTCTTTGACGGGCAAGGACGTCCTGAGGCACCCCACCTGG 1845

Qy   481 CCCCGGAAGAGCGTGTGGGACACGGCTCCGACCCCCAGCCGGCGGCGCCGCCTGACGGACAGCTAC 540
          |||| |||||| || | | ||| |||||||||| ||| |||||||||||||||||||||
Db   1846 CCCCAGAAGAGCCGTGTGGCAtGGCTGCATGGCTTCGGCGGCTCGGACCCCAACCGGCGCCAACCGGCCTGACGGCTGACGAGCTAC 1905

                                                                                       Qy   541 TGCGGAGACGGTGGCGACGAGGCCCCGGCGGCGCACCGGCGGCAGGCGTCGTCGCTGCTGGCG 600
          || |||||||||||||||| ||| || | ||||||||| ||| ||||||||||||||| |
Db   1906 TGTGAGACGGTGGCGACGGAGGCTCCCTCGGCCACGGGCCAGGCCTCCTCGCTGCTGGGG 1965

Qy   601 GGCAGGCTGCTGGAGCAGGAGGCCGGAGTGCCGGAGCTGCCACGCGCCTTCGTGGTGCTCTTGCATC 660
          |||||||| ||||| |||| ||| | |||||| ||||||||| | ||| |||| | ||| ||||||
Db   1966 GGCAGGCTCCTGGGCGAGAGTGCGGCGAGCTGCCATCACGGCCCATCATCGTGTCTGCATT 2025

Qy   661 GAGAACAGCGTCATGACCTTCCTTCTCCAAGTAGGGCCCGCGCGCCCACGGACGAGGCGGGG 720
          ||||||||| ||||||| | ||| ||||||||| |||| ||||| ||| | |||| |||
Db   2026 GAGAACAGCTTCATGACTGCCTCCAAGTAGCCACCGCCGCCTGGATGGCCGGAAGAGG 2085

Qy   721 GAGGGGGCGCCCGCAGGAGGAGCATCCGCCGCCCGGGG----------------- 756
          || | || | | |||| || ||| | || || ||||
Db   2086 ACCGGCGGCGTCGGAGGAGCCCCACCGTGGCCGAGGGAGCAGGCGCAGCCCCTGCCC 2145

Qy   757 --GGGCCTGGCCGGGACCGCTTGCCTGCCACCG-TCACGTTTAATGTAATCCTCAAGAAATA 813
            |||| ||||| || || | |||| || |||| |||||||| ||| | ||||||||||||||
Db   2146 CAGGACCCTGGCTGCCCATACTTTCCTGTATAGTTCACGTTTCATGTTCATGTAATCCTCAAGAAATA 2205

Qy   814 AAAGGAAGCCAAAGAG 829
          ||||||||||||||||
Db   2206 AAAGGAAGCCAAAGAG 2221

RESULT 7
AAX78379
ID   AAX78379 standard; cDNA; 3394 BP.
XX
AC   AAX78379;
XX
DT   25-AUG-1999 (first entry)
XX
DE   Human alpha1 (XVIII) collagen cDNA.
XX
KW   Alpha1(XVIII) collagen; mimetic; endostatin; atomic coordinate; library;
KW   anti-angiogenic; heparin binding domain; receptor binding domain; mimic;
KW   alpha-helix A domain; carbohydrate recognition domain; CRD domain;
KW   treatment; angiogenesis; tumour; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..2055
FT   /tag= a
FT   /product= "alpha1(XVIII) collagen"
FT   /note= "Partial sequence, no start codon given"
XX
PN   WO9931616-A1.
XX
PD   24-JUN-1999.
XX
PF   16-DEC-1998;   98WO-US026783.
XX
PR   16-DEC-1997;   97US-0069727P.
XX
PA   (HARD) HARVARD COLLEGE.
XX
PI   Olsen BR, Hohenester E, Timpl R, Sasaki T;
XX
DR   WPI; 1999-395243/33.
DR   P-PSDB; AAY25113.
XX
PT   Identifying mimetics of mammalian endostatin.
XX
PS   Disclosure; Fig 5A-C; 75pp; English.
XX
CC   This invention describes a novel method for identifying mimetics of
CC   mammalian endostatin. The method comprises identifying a compound having
CC   atomic coordinates with non-trivial similarity to selected coordinates of
CC   atoms of a mammalian endostatin involves (a) providing a library of
CC   atomic coordinates of compounds in a library of candidate compounds, (b)

CC   comparing the library of atomic coordinates to the selected coordinates
CC   of a mammalian endostatin and (c) selecting from the library at least one
CC   candidate compound on the basis of selection criteria which include
CC   similarities between the atomic coordinates of the selected candidate
CC   compound and the atomic coordinates of the mammalian endostatin. The
CC   invention also describes the use of an anti-angiogenic fragment of
CC   endostatin comprising a domain selected from a heparin binding domain, a
CC   receptor binding domain, and exposed on alpha-helix A domain, and a
CC   carbohydrate recognition domain (CRD) domain. The methods can be used for
CC   designing and selecting endostatin mimics. The compounds identified can
CC   be used for treating undesired angiogenesis, e.g. tumours. This sequence
CC   encodes human alpha1(XVIII) collagen which is used in the description of
CC   the method
XX
SQ   Sequence 3394 BP; 668 A; 1130 C; 1080 G; 516 T; 0 U; 0 Other;

Query Match   59.9%; Score 496.8; DB 2; Length 3394;
Best Local Similarity   77.1%; Pred. No. 2.1e-72;
Matches 660; Conservative   0; Mismatches 157; Indels 39; Gaps   3;

Qy   1 CCCTGGCGGGCAGAGATGACATCCTGGCCGGCCCCCCCGGCCCTGCTGGACACCCCAGCCCTAC 60
Db   1378 CCCTGGCGGGCAGAGATGACATCCTGGCCGGCCCCCCCGGCCCTGCTGGACACCCCAGCCCTAC 1437
Qy   61 CCCGGGGGCCCCCGCCACCACGGCTCCTACGTGACTTCCAGCCGGCGTCGCCCCACTGGTGTGGG 120
Db   1438 CCCGGAGGCCCGGCCGCACCACAGCTTCTACGTGCACTGGCCGCCCGGCAGACCCACAAG---- 1493
Qy   121 CCCGTCCACACCCACCACCACCACCAGAGTTCCAGCTGGTGCTGCACCTGGCC 180
Db   1494 -------CCCACCCGCCCAGAGCCACCGCGACTTCCAGCCGCGTGCTGCACCTGTTGCG 1545
Qy   181 CTGAACAGCCCGGAGCCCGGCGGCATGGGAGGCATCCGGGAGGCGGGACTTCCAGTGCTTC 240
Db   1546 CTCAACAGCCCCCGTCGTCAGGCGGCATGCGGGCATCCGGGCGGGGCCGACTTCCAGTGCTTC 1605
Qy   241 CAGCAGGCGGCGCGGCCGGCGGGGTGGCCCGACCTTCCGGGCCTTCCTGTGTCGCGGCTG 300
Db   1606 CAGCAGGCGGCGGCCGGCCGTGGGGCGGCTGGCGGGGCGCACCTTCCGCGCCTTCCGTGTCCTCGCGGCCTG 1665
Qy   301 CAGGACCCTCTACAGCATCGTGCCGCCGGCCGGACCGGCTGGCCGTGCCCGTCGTCAACCTC 360
Db   1666 CAGGACCTTGTACAGCATCGTGCCGCCGGCCGGGACCGGCTGGCCGTGCCCATCGTCAACCTC 1725
Qy   361 AGGACGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTCGGGCTCCGGAGGGGCCAGCTG 420
Db   1726 AAGGACGAGAGCTGCTGTTTCCCAGCTGGGAGGCTTCGTTCTCCAGGGCCTGAGGGGTCCGGTG 1785
Qy   421 AAGCCCGGGGCCCGATCTTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCCTGG 480
Db   1786 AAGCCCGGGGGACGCAGCATCTTCTCCTTTGACGGCAAGGACGTCCTGAGGCACCCCACCTGG 1845
Qy   481 CCCCGGAGAGGCGTGTGGCACGGCTCCACACCCAGCCGGGCGCGCCTGACCAGCACGACAGCTAC 540
Db   1846 CCCCAGAAGAGCGTGTGGCATGGCCATGCGGGAGCTGGATCGGACCTCGACCCAACGGGCAGCTGACCGAGCTAC 1905
Qy   541 TGCGAGACGTGGCGACGGACGAGGCCCGGCCGGCAGGCCCAGGGCTGTCGTGCTGCTGGCG 600
Db   1906 TGTGAGACGTGGCGGACGGACCGGACGAGGCCCTCGGCCCGGCGCGGGCCGCCTCTCCGCTGCTGTCGGGG 1965
Qy   601 GGCAGCTGGCTGCTGGAGCGGCAGGAGGGGCCGGAGCTGCCGCCACGCCCTTCGTGGTTGCTCTGCATC 660
Db   1966 GGCGAGGCTCCTGGGGCAGAGTGCCGGGCAGAGTGCCGAGCTGCCATCACGCGCTACATCGTGCTCTGCATT 2025
Qy   661 GAGAACAGCCTCATGACCTCCTTCTCCAAGTAGGGGCCCGCGGCCCACGGACAGGCGGGG 720
Db   2026 GAGAACAGCTTCATGAGCTTCATGACTGACTAGCCACCGGCCGATGCAGATGCGGAGAGG 2085
Qy   721 GAGGGGGCGCGCCCGCAGGAGCATCCGCCGCCCGGGG--------- 756
Db   2086 ACCGGCGGCGCTCGGAGGAAGCCCCCACCGTGGGCACGTGGGAGGCAGCCGGCAGCCCCTGGCCC 2145
Qy   757 --GGGGCTGCCTGCTGGCGGGAGCGGCGCGTTGCCTGCACCG-TCACGTTTAATGTAATCCTCAAGAAATA 813
Db   ... (continued)

Db   2146 CAGGACCTGGCTGCCATACTTCCTGTATAGTTCACGTTCATGTAATCCTCAAGAAATA 2205
Qy   814 AAAGGAAGCCAAAGAG 829
Db   2206 AAAGGAAGCCAAAGAG 2221

RESULT 8
ABN95680
ID   ABN95680 standard; DNA; 3394 BP.
XX
AC   ABN95680;
XX
DT   13-AUG-2002 (first entry)
XX
DE   Gene #2178 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour; cytostatic; expression profile; disease state;
KW   disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS   Homo sapiens.
XX
PN   WO200229103-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US030589.
XX
PR   02-OCT-2000; 2000US-0237054P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
DR   WPI; 2002-426119/45.
XX
PT   Diagnosing and detecting the progression of liver cancer, hepatocellular
PT   carcinoma or metastatic liver tumor in a patient, involves detecting the
PT   level of expression of two or more genes in a liver tissue sample.
XX
PS   Claim 1; SEQ ID NO 2178; 298pp; English.
XX
CC   The invention relates to a novel method for diagnosing and detecting the
CC   progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC   tumour in a patient, and differentiating metastatic liver cancer from
CC   hepatocellular carcinoma in a patient, involving detecting the level of
CC   expression of two or more genes represented in ABN93503-ABN97455 in a
CC   tissue sample. The method of the invention has hepatotropic, and
CC   cytostatic activity. The method is useful for diagnosing and detecting
CC   the progression of liver cancer, hepatocellular carcinoma and metastatic
CC   liver carcinoma in a patient. The method is useful for identifying
CC   expression profiles which serve as useful diagnostic markers as well as
CC   markers that can be used to monitor disease states, disease progression,
CC   drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3394 BP; 668 A; 1130 C; 1080 G; 516 T; 0 U; 0 Other;

Query Match   59.9%; Score 496.8; DB 6; Length 3394;
Best Local Similarity   77.1%; Pred. No. 2.1e-72;
Matches 660; Conservative   0; Mismatches 157; Indels 39; Gaps   3;

Qy   1 CCCTGGCGGGCAGAGATGACATCCTGGCCGGCCCCCCCGGCCCTGCTGGACACCCCAGCCCTAC 60
Db   1378 CCCTGGCGGGCAGAGATGACATCCTGGCCGGCCCCCCCGGCCCTGCTGGACACCCCAGCCCTAC 1437
Qy   61 CCCGGGGGCCCCCGCCACCACGGCTCCTACGTGACTTCCAGCCGGCGTCGCCCCACTGGTGTGGG 120
Db   1438 CCCGGAGGCCCGGCCGCACCACAGCTTCTACGTGCACTGGCCGCCCGGCAGACCCACAAG---- 1493

```
Qy   121 CCCGTCCACACCCACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCC 180
             ||||||        ||||||||||||||      |||||||||||||  ||
Db  1494 ------CCCACCCGCCCACAGCCCACGGACTTCCAGCCTCCACCTGGTTGCG 1545

Qy   181 CTGACAAGCCCGCAGCCGGGCGGCGATCGGGAGGCATCCGGGGAGCCGACTTCCAGTGCTTC 240
             ||        ||||||||||||||  || |||||   ||||||||||||||||||||
Db  1546 CTCACACAGCCCCCGTGTCAGGGGGCATGCGGGCGGCCGCATCCGGGGCCGACTTCCAGTGCTTC 1605

Qy   241 CAGCAGGCGCGCGGCGGGCGCGGGGTGGCCGGGCACCTTCCTGTCGTCGTGCGGCGGCCTG 300
         ||||||||||||||||||||||||||||||||||||||||||    ||||  ||||||||
Db  1606 CAGCAGGCGCGCGGCGGGCGCGGGGTGGCCGGGCACCTTCCTTCGTCCTCGTGCGGCGCCCTG 1665

Qy   301 CAGGACCCTACAGCATCGTGCCGCCGACGCGCACCGGGGTGCCCGTCGTCAACCTC 360
         ||||||||||||||||||  |||  ||||||||  |||||||||   ||| |||||
Db  1666 CAGGACCCTGTACAGCATCGTGCCGCCGACGCGCACCGGGGTGCCCGTCGTCAACCTC 1725

Qy   361 AGGAGCAGGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTG 420
         ||||  ||||||||||||||||||||||||||||||  ||||  ||||||||||||||||||
Db  1726 AAGGACGAGCTGCTGTTTCCCAGCTGGGAGGCTCTGTTCTCCAGGGTCTGAGGGGTCCGCTG 1785

Qy   421 AAGCCCGGGGCCCGCATCTTCTTCTTGACGGCAGAGATGTCCAGCACCCCGCCTGG 480
         ||||||||||||||||||||||    |||||||||||||||||||||||  |||||
Db  1786 AAGCCCGGGGCACGCATCTTCTTTGACGGCAAGGACGTGCTCCTGAGGCACCCACCTGG 1845

Qy   481 CCCCGGAAGAGCGTGTGGCACGGCTCCGACCCCAGCGGGCGCGGGCCTGACCGACAGCTAC 540
         ||||  |||||||||||  ||||||||||  |||||||||||||||||||||||||||||
Db  1846 CCCCAGAAGAGCGTGTGGCATGCGGCCCCAACGGGGCGCAGGCTGACCGACGAGCTAC 1905

Qy   541 TGCGAGACGTGGGCGGACAGGAGGCCCCGGGCCGCCACCGGCGTCGTTCGCTGCTGGCG 600
         |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db  1906 TGTGAGACGTGGGCGGACAGGAGGCCCCGGGCCGCCACCGGCTCCTCGCTGCTCTGGGGG 1965

Qy   601 GGCAGGCTGCTCGAGCAGGAGGCGCCGCAGCGTCGCCCACGCCCTTCGTGGTGTCTCGCATC 660
         |||||||||||||||||||||||||||||||||||||||    |||||||||||||||||
Db  1966 GGCAGGCTCCTCGGGGCAGAGTGCCGGCAGGTGCCATCACCGCCTACCATCGTGCTCTGCATT 2025

Qy   661 GAGAACCAGCCGTCATGACCTCCTTCTCCAAGTAGGGCGGCCCACGGACAGGCGGGGG 720
         ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db  2026 GAGAACAGCTTCATGACTGACTAGCCACCGCCTGGATGCAGAGTGCACCGGCCCGAGAGAG 2085

Qy   721 GAGGGGGCCCGCCCAGAGCATCCGCCGCCCCGGGG------------ 756
         ||||||||||||||||||||||||||||||||||||
Db  2086 ACCGGCCGCCCCGGAGGAAGCCCCACCGTGGGCAGGGAGCGACGCCGGCCAGCCCCTGGCCC 2145

Qy   757 --GGGCCTGGCCGGGGAGCGCGCTTGCCTGCACCG-TCACGTTTAATGTAATCCTCAAGAAATA 813
              |||||||| ||||||||||    |||||||||||||||||||||||||||||||||||
Db  2146 CAGGACCTGGCTGCCATACTTTCCTGTATAGTTCACGTTTCATGTAATCCTCAAGAAATA 2205

Qy   814 AAAGGAAGCCAAAGAG 829
         ||||||||||||||||
Db  2206 AAAGGAAGCCAAAGAG 2221
```

RESULT 9
ABK09977
ID   ABK09977 standard; DNA; 900 BP.
XX
AC   ABK09977;
XX
DT   21-MAY-2002 (first entry)
XX
DE   Synthetic plasmid pEnd-HR#1 FPD fusion DNA sequence.
XX
KW   Mouse; Ig signal peptide; mIgSP; functional protein domain; FPD;
KW   primary translational product; PTP; DNA construct; regulatory DNA;
KW   DNA targeting segment; regulatory factor; single regulatory unit;
KW   monoclonal antibody; recombination-derived alteration; blood product;
KW   human; COL18A1; gene; mutant; fusion protein; ds.
XX
OS   Mus sp.
OS   Homo sapiens.
OS   Synthetic.
OS   Chimeric.
XX
FH   Key   Location/Qualifiers
FT   exon   1..117
FT   /tag= a
FT   /number=1
FT   /note="Mouse Ig signal peptide (mIgSP) exon"
FT   CDS   61..888
FT   /tag= b
FT   /product= "pEnd-HR#1 FPD fusion protein"
FT   sig_peptide   61..117
FT   /tag= c
FT   /note="Mouse Ig signal peptide (mIgSP)"
FT   mat_peptide   118..885
FT   /tag= d
FT   /note="Human COL18A1 protein exons 38-41"
FT   exon   118..363
FT   /tag= e
FT   /number= 2
FT   /note="Human COL18A1 protein exon 38"
FT   exon   364..558
FT   /tag= f
FT   /number= 3
FT   /note="Human COL18A1 exon 39"
FT   exon   559..677
FT   /tag= g
FT   /number= 4
FT   /note="Human COL18A1 exon 40"
FT   exon   678..885
FT   /tag= h
FT   /number= 5
FT   /note="Human COL18A1 exon 41"
XX
PN   WO200210372-A1.
XX
PD   07-FEB-2002.
XX
PF   01-AUG-2001; 2001WO-GB003455.
XX
PR   01-AUG-2000; 2000GB-00018876.
XX
PA   (ISTF) ARS APPLIED RES SYSTEMS HOLDING NV.
PA   (CHAP/) CHAPMAN P W.
XX
PI   Chapman PW, De Luca G, Falciola L;
XX
DR   WPI; 2002-195963/25.
DR   P-PSDB; AAU76689.
XX
PT   Producing functional protein domain by growing host cell transfected with
PT   DNA construct having regulatory DNA and DNA targeting segment, and
PT   optionally culturing homologously recombinant cell and collecting
PT   protein.
XX
PS   Example; Fig 8; 116pp; English.
XX
CC   The present invention relates to a new method of producing a protein,
CC   such as functional protein domain, that is either C- or N-terminus of the
CC   primary translational product (PTP) of a gene, where the protein has
CC   biological activity which is distinct from PTP. The method of the
CC   invention involves growing a host cell transfected with a DNA construct
CC   comprising a regulatory DNA and a DNA targeting segment. This method is
CC   useful for producing a functional protein domain of proteins such as
CC   regulatory factors, blood products and monoclonal antibodies. The method
CC   described in the invention allows controlled and precise modification of
CC   the host cell genome in order to produce functional protein domain (FPD).
CC   The amount of exogenous sequence to be integrated in the host cell genome
CC   is very limited since, as coding sequence, the original coding sequence
CC   present in the host cell genome itself is used. Use of the host cell
CC   sequence encoding FPD also provides the advantages of both eliminating
CC   any recombination-derived alteration of such coding sequence, and also
CC   making use of the same post-transcriptional (e.g., splicing) and/or post-
CC   translational (e.g. glycosylation, phosphorylation) processes that are

CC   actually applied in vivo for the maturation of FPD. The use of a single
CC   regulatory unit eliminates the necessity of manipulating the
CC   complementary DNA coding for the PTP to isolate the segment coding for
CC   the FPD, and adapt it to the expression vector. The present nucleic acid
CC   sequence encodes the plasmid pEnd-HR#1 FPD fusion protein of the
CC   invention. The FPD fusion protein contains the mouse Ig signal peptide
CC   (mIgSP) sequence fused to exons 38-41 of the human COl18Al sequence
XX
SQ   Sequence 900 BP; 146 A; 336 C; 274 G; 144 T; 0 U; 0 Other;

Query Match   59.5%;   Score 493.2;   DB 6;   Length 900;
Best Local Similarity   83.9%;   Pred. No. 8.9e-72;
Matches 574;   Conservative   0;   Mismatches   98;   Indels   12;   Gaps   1;

Qy   1 CCCTGGCGGGCAGATGACATCCTGGCGCGGCCCCCGCCGCTGACCCCAGCCCTAC 60
           |||||||||||||||||||||||||||||||||||| ||||| ||||||||||
Db   211 CCCTGGCGGGCGAGTGACATCCTGGCGCAGCCCCCT CGCCTGCGCCCAGCCCTAC 270

Qy   61 CCCGGGGCCCCGACCACAGGCTCCTACGTGCACTTCCAGCCGGCTCGCCCACTGGTGGG 120
           ||||||||| ||||||||||| ||||||||||||||||||||||
Db   271 CCCGGAGCCCCGACCACAGCTCCTACGTGCACCTGGCGCCGGCGGGACCCCACCACAAG--- 326

Qy   121 CCGTCCACACCCACCACCACCACCAGGACTTCAGCTGGTGCTGCACCTGGTGGCC 180
           |||||||||||| ||||||||||||||||||||||||||||||||||
Db   327 -------CCACCCGCCCACAGCCACCACCGCACCCGACCCGTGCTTCCACCTGGTTGCG 378

Qy   181 CTGAACAGCCCGCAGCCCGGCGGCCATGCGGCGGCCATCCGGGAGCGGACTTCCAGTGCTTC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   379 CTCAACAGCCCCGTGTCAGGCGGCATCGCGGGGCCGACTTCCAGTGCTTC 438

Qy   241 CAGCAGGCCGGCGGCCGGCGGGGCTGCGCCGGGGCCCTTCCTGTCGTCGCGGCTG 300
           ||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   439 CAGCAGGCCGGCGGCCGGCGTGGGGCTGGGGCGGCACCTTCCGGCCTTCGTCCTCCGCGCCTG 498

Qy   301 CAGGACCTCTACAGGATGTGCGCCGGACCCCGCACCGGGGTGCCCGTCGTCAACCTC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   499 CAGGACCTGTACAGCATCGTGCGCCGTGCCCAGCCCGTGCCCATCGTCAACCTC 558

Qy   361 AGGGAGGAGTGCTGCTTCCCCAGCTGGGAGGGCCTTATTCTCGGGCTCGAGGGCCAGCTG 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   559 AAGGACGAGCTGCTGTTTCCCAGCTGGGAGGGCTCTGTTTCTCAGGCTCTGAGGGTCCGCTG 618

Qy   421 AAGCCCGGGGCCCGGGCCCGATCTTCTTCTTCTTTCGACGGGCAGAGATGTCCTGCAGCAGAGATGTCCTGCACCCCGCCTGG 480
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   619 AAGCCCGGGGCACGGGCCACGCATCTTCTTCTTTGACGGGCAGAGCGTCCCAGGCGCACCCCACTGG 678

Qy   481 CCCCCGAAGAGAGCGTGTGGCACGGTCCGGCTCCCCGCGCGGCCCAGCCGACCAGCTAC 540
           ||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   679 CCCCAGAAGAGAGCGTCTGGCATGGCTCGGACGAGCCTGGGGCGCAGGCGGAGAGCTAC 738

Qy   541 TGCGACAGGTGGCGGACGGAGCGCCCCGGCGGGCCACCGGGCAGGCGTCGTCGCTGGCG 600
           |||||||||||||| |||||||||||| |||||||||||||||||||||||
Db   739 TGTGAGACGTGGCGGACCGAGCGGCGCTCCCTGGCCGCACGGGGCCTCCCGCTGGGGG 798

Qy   601 GGCAGGCGCTGCTGAGGCAGGAGGCGGCGAGGAGCGGCCTTCGTGGTGCTCTGCATC 660
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   799 GGCAGGCGGCTCCTGGGCAGAGTGCGCGGCGAGCTGCCGACTGCCATCACGTGCTGCTTGCATT 858

Qy   661 GAGAACAGGCGTCATGACCTCCTTC 684
           ||||||||||||||||||||||||
Db   859 GAGAACAGCTTCATGACTGCCTCC 882

RESULT 10
ABN85301
ID   ABN85301 standard; DNA; 4551 BP.
XX
AC   ABN85301;
XX
DT   30-SEP-2002 (first entry)
XX
DE   Human collagen XVIII coding sequence.
XX

KW   Human; antirheumatic; antiarthritic; gene therapy; anti-angiogenic;
KW   rheumatoid arthritis; collagen; endostatin; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..4551
FT   /*tag= a
FT   /product= "Human collagen XVIII"
XX
PN   WO20253191-A1.
XX
PD   11-JUL-2002.
XX
PF   03-JAN-2002; 2002WO-KR000001.
XX
PR   05-JAN-2001; 2001KR-00000691.
XX
PA   (VIRO-) VIROMED LTD.
XX
PI   Kim J, Ho S, Park E, Kim S;
XX
DR   WPI; 2002-583596/62.
DR   P-PSDB; ABB83471.
XX
PT   Novel composition for gene therapy against rheumatoid arthritis,
PT   comprising a DNA encoding anti-angiogenic protein or its parts.
XX
PS   Disclosure; Page 60-70; 84pp; English.
XX
CC   The present invention relates to a composition for gene therapy,
CC   comprising a DNA encoding an anti-angiogenic protein, which shows
CC   therapeutic effects on rheumatoid arthritis. The composition is useful
CC   for treating rheumatoid arthritis and the gene therapy is effective,
CC   lasting for 14 days. The present sequence is the coding sequence for
CC   human collagen XVIII. Endostatin, which consists of the C-terminal 183
CC   residues of collagen XVIII, was used as an anti-angiogenic protein
XX
SQ   Sequence 4551 BP; 780 A; 1597 C; 1522 G; 652 T; 0 U; 0 Other;

Query Match   59.5%;   Score 493.2;   DB 6;   Length 4551;
Best Local Similarity   83.9%;   Pred. No. 7.8e-72;
Matches 574;   Conservative   0;   Mismatches   98;   Indels   12;   Gaps   1;

Qy   1 CCCTGGCGGGCAGATGACATCCTGGCGCGGCCCCCGCCGCTGACCCCAGCCCTAC 60
           |||||||||||||||||||||||||||||||||||| ||||| ||||||||||
Db   3874 CCCTGGCGGGCGAGTGACATCCTGGCGCAGCCCCCT CGCCTGCGCCCAGCCCTAC 3933

Qy   61 CCCGGGGCCCCGACCACAGGCTCCTACGTGCACTTCCAGCCGGCTCGCCCACTGGTGGG 120
           ||||||||| ||||||||||| ||||||||||||||||||||||
Db   3934 CCCGGAGCCCCGACCACAGCTCCTACGTGCACCTGGCGCCGGCGGGACCCCACCACAAG--- 3989

Qy   121 CCGTCCACACCCACCACCACCACCAGGACTTCAGCTGGTGCTGCACCTGGTGGCC 180
           |||||||||||| ||||||||||||||||||||||||||||||||||
Db   3990 -------CCACCCGCCCACAGCCACCACCGCACCCGACCCGTGCTTCCACCTGGTTGCG 4041

Qy   181 CTGAACAGCCCGCAGCCCGGCGGCCATGCGGCGGCCATCCGGGAGCGGACTTCCAGTGCTTC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4042 CTCAACAGCCCCGTGTCAGGCGGCATCGCGGGGCCGACTTCCAGTGCTTC 4101

Qy   241 CAGCAGGCCGGCGGCCGGCGGGGCTGCGCCGGGGCCCTTCCTGTCGTCGCGGCTG 300
           ||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   4102 CAGCAGGCCGGCGGCCGGCGTGGGGCTGGGGCGGCACCTTCCGGCCTTCGTCCTCCGCGCCTG 4161

Qy   301 CAGGACCTCTACAGCATGCGTGCGCCGGACCCCGCACCGGGGTGCCCGTCGTCAACCTC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4162 CAGGACCTGTACAGCATCGTGCGCCGTGCCCAGCCCGTGCCCATCGTCAACCTC 4221

Qy   361 AGGGACGAGGTGCTTCTGTTTCCCAGCTGGGAGGGCCTTATTCTCGGGCTCTGAGGGTCCGCTG 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4222 AAGGACGAGCTGCTGTTTCCCAGCTGGGAGGGCTCTGTTTCTCAGGCTCTGAGGGTCCGCTG 4281

Qy   421 AAGCCCGGGGCCCGGGCCCGCATCTTCTTCTTTCGACGGGCCCAGATGTCCTGCAGCAGAGATGTCCTGCAGCACCCCGCCTGG 480

CC   cells expressing the polypeptide are useful in treating pancreatic cancer
CC   and stimulating an immune response. The polynucleotides can be used as
CC   probes or primers for nucleic acid hybridisation, in the design and
CC   preparation of ribozyme molecules for inhibiting expression of the tumour
CC   polypeptides and proteins in the tumour cells, in vaccines and for gene
CC   therapy. Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4551 BP; 780 A; 1597 C; 1522 G; 652 T; 0 U; 0 Other;

Query Match   59.5%;   Score 493.2;   DB 6;   Length 4551;
Best Local Similarity   83.9%;   Pred. No. 7.8e-72;
Matches 574;   Conservative 0;   Mismatches 98;   Indels 12;   Gaps 1;

Qy   1 CCCTGCGGGCGAGATGACATCCTGCCGGCCCCGCGGCTGGTGCACTTCCAGCCCTAC 60
          |||||||||||||||| |||||||||||||||||||| |||||||||||||||||||
Db   3874 CCCTGCGGGCCAGATGACATCCTGCCAGCCCCTGGCTGGTGCACTTCCAGCCCTAC 3933

Qy   61 CCCGGGGCCCCGCACCACCGCTCCTACGTGCACTTCAGCCGGCTGCCCACTGGTGGGG 120
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||||---
Db   3934 CCCGGAGCCCCGCACCACCAGCTCCTACGTGCACCTGCCCGGCGGGCGACCCACCAAG--- 3989

Qy   121 CCCGTCCACACCACCACCACCACCAGGACTTCAGGACTTCCAGCCCGGTGCTGCACCTGGTGGCC 180
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   3990 --------CCCACCCGGGCCACAGCCACCCGCAGTTCCAGCCCGGTGCCCACCTGGGTTCG 4041

Qy   181 CTGAACAGCCCCGCAGCCGGCGGCATGGCAGGCATCCGGGGAGCGACTTCCAGTGCTTC 240
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   4042 CTCAACAGCCCCCTGTCAGCGGCATGGGGCATCCGGGGCCGGACTTCCAGTGCTTC 4101

Qy   241 CAGCAGGCGGCGCGCCGGGTGGCCCGGCACCTTCGGCCCTTCCTGTCGTCGCGGCTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4102 CAGCAGGCGGCGGCGCCGTGGGGCTGGCGGGCACCTTCGGCCCTTCGTCCTCGGCCGGCTG 4161

Qy   301 CAGGACCTCTACAGACATGTGGCCCGCCGCCCGACCGGGGTGCCCGTGCGTCAACCTC 360
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   4162 CAGGACCTGTACAGACATCGTGCCCGCCGTGCCAGCCGCGGGCCCCATCGTCAACCTC 4221

Qy   361 AGGGACGAGGTGCTCTTCCCAGCTGGCGAGGGCCTTATTCTGGGTCCGAGGGCCAGCTG 420
          |||||||||||||||||||||||||||||||| || ||||||| |||||||||||||
Db   4222 AAGGACGAGCTGCTGTTTCCCAGCTGGGAGGGCTCGTGTCTGTTTCTCAAGGCTGGAGGGTCCGCTG 4281

Qy   421 AAGCCCGGGCCCGCCATCTTCTTTCTGCACGGGCAGAGATGTCCTGCAGCAGATGCTTCTGCTGGG 480
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   4282 AAGCCCGGGCCGCCACCGCATCTTCTCTTTGACGGCCAAGGACGTTCCTGAGGCCACCTGG 4341

Qy   481 CCCCGAAGAGCGTGTGGCACGGCTCCACCCCAGCCGGCGCCCAGCCCGCTGACCGACAGCTAC 540
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   4342 CCCCAGAAGAGCGTGTTGGCATGGCTCCGGCACGCCCCAACGGGCCGCCAGGGCGAGAGCTAC 4401

Qy   541 TGCGAGACGTGGCGACGGAGCCCCGGGGGCCCCGGCAGGCGTCGTCGCTGCTGGCG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4402 TGTGAGACGTGGCGACGGAGCCCGGGGCGCCGGCAGGCCTTCCCTGCTGCTGGGG 4461

Qy   601 GGCAGGCTGCTGTGAGCAGGAGGCCGCAGAGGCCCACGCGCCTTCGTGGTGCTCATC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4462 GGCAGGCTCCTGGGCGAGAGTGCCGCGAGTGCCGCGAGCTGCCATCACGCGCTACATCGTGCATT 4521

Qy   661 GAGAACAGGGTCATGACCTCCTTC 684
          |||||||||||||||||||||
Db   4522 GAGAACAGCTTCATGACTGCCTCC 4545

RESULT 11
ABV94763
ID   ABV94763 standard; cDNA; 4551 BP.
XX
AC   ABV94763;
XX
DT   14-JAN-2003   (first entry)
XX
DE   Human pancreatic cancer expressed cDNA SEQ ID NO 144.
XX
KW   Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;
KW   cytostatic; tumour; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200260317-A2.
XX
PD   08-AUG-2002.
XX
PF   30-JAN-2002; 2002WO-US002781.
XX
PR   30-JAN-2001; 2001US-0255305P.
PR   31-JAN-2001; 2001US-0265682P.
PR   09-FEB-2001; 2001US-0267568P.
PR   28-MAR-2001; 2001US-0278651P.
PR   28-APR-2001; 2001US-0287112P.
PR   16-MAY-2001; 2001US-0291631P.
PR   12-JUL-2001; 2001US-0305484P.
PR   20-AUG-2001; 2001US-0313999P.
PR   27-NOV-2001; 2001US-0333626P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Benson DR, Kalos MJ, Lodes MJ, Persing DH, Hepler WT, Jiang Y;
XX
DR   WPI; 2002-627435/67.
DR   P-PSDB; ABP68617.
XX
PT   New isolated polynucleotide and pancreatic tumor polypeptides, useful for
PT   diagnosing, preventing and/or treating cancer, particularly pancreatic
PT   cancer.
XX
PS   Claim 1; SEQ ID NO 144; 300pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising: (a)
CC   any of a group of over 4000 nucleotide sequences (ABV94628-ABV99145); (b)
CC   complements of (a); (c) sequences consisting of at least 20 contiguous
CC   residues of (a); (d) sequences that hybridize to (a), under moderately
CC   stringent conditions; (e) sequences having at least 75% or 90% identity
CC   to (a); or (f) degenerate variants of (a). Polypeptides (ABP68596-
CC   ABP68637) encoded by (I) and oligonucleotide can be used to detect cancer
CC   in a patient and compositions comprising polypeptides, polynucleotides,
CC   antibodies, fusion proteins, T cell populations and antigen presenting

RESULT 12
AAC62025
ID   AAC62025 standard; DNA; 968 BP.
XX
AC   AAC62025;
XX
DT   06-MAR-2001   (first entry)
XX
DE   Nucleotide sequence of vaa-endostatin fusion protein in pANT3052.

XX
KW   Streptomyces sp. strain C5; SnpA; S. venezuelae; alpha-amylase;
KW   endostatin; cancer; tumour growth; angiogenesis; ss.
XX
OS   Synthetic.
OS   Streptomyces sp.
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   misc_binding 220..234
FT   /*tag= a
FT   /note= "proposed SnpR binding site"
FT   RBS   307..311
FT   /*tag= b
FT   CDS   318..968
FT   /*tag= c
FT   sig_peptide 318..401
FT   /*tag= d
FT   /note= "vaa signal sequenvce"
FT   mat_peptide 402..968
FT   /*tag= e
FT   /note= "endostatin"
XX
DR   WO200060945-A1.
XX
PN   WO20000060945-A1.
XX
PD   19-OCT-2000.
XX
PF   12-APR-2000; 2000WO-US009747.
XX
PR   13-APR-1999; 99US-0129084P.
XX
PA   (MERI) MERCK & CO INC.
XX
PI   Desanti CL, Strohl WR;
XX
DR   WPI; 2000-686970/67.
DR   P-PSDB; AAB30495.
XX
PT   Preparation of soluble recombinant endostatin involves transforming
PT   Streptomycete host with expression vector comprising nucleotide sequence
PT   encoding endostatin operably linked to linker and leader peptide.
XX
PS   Example 1; Fig 10A-B; 57pp; English.
XX
CC   The present sequence encodes a fusion protein of vaa and endostatin. The
CC   specification describes a method for the production of soluble,
CC   recombinant human endostatin in Streptomyces. Leader sequences of
CC   Streptomyces sp. strain C5 SnpA and S. venezuelae alpha-amylase proteins
CC   are linked to the N-terminal of endostatin. This ensures that endostatin
CC   protein is produced as a secreted, soluble protein which needs no
CC   refolding, is stable in the fermentation broth and is produced in large
CC   quantities. The method is used for preparing soluble recombinant human,
CC   murine or primate endostatin, which is useful in the treatment of cancer,
CC   inhibition of tumour growth, inhibition of angiogenesis, isolation of
CC   receptors for endostatin and for identification of anti-angiogenic
CC   compounds in assays. The endostatin protein is produced as a secreted,
CC   soluble protein which needs no refolding, is stable in the fermentation
CC   broth and is produced in large quantities. Streptomycetes are amenable
CC   for cultivation in large fermentations allowing for large quantities of
CC   soluble endostatin to be produced
XX
SQ   Sequence 968 BP; 153 A; 335 C; 331 G; 149 T; 0 U; 0 Other;

Query Match   51.3%;   Score 425.6;   DB 3;   Length 968;
Best Local Similarity   83.0%;   Pred. No. 9.1e-6;;
Matches 485;   Conservative 0;   Mismatches 99;   Indels 0;   Gaps 0;

Qy   101 CGGCTCGCCCACTGGTGGCCCGTCCACACCCACACCCACACCCACCAGGACTTCCAGC 160
          |||||||||||||||||| ||||||||||||||||||| |||||||||||||| |||||||
Db   379 CGGGCAACGCCCCGCCGCGCAGGCCGTCCCGCCCCGGATCCGCAGCCACCGCCGGACTTCCAGC 438

Qy   161 TGGTGCTGCACCTGGTGGCCCTGAACAGCCCGCCGGCCGGGCGGCCATCGAGGCATCCGGG 220
          |||||||| | ||||||| || |||||| | ||||||| || |||||||||||||||

Db   439 CGGTGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCGCAGCGCATGCGGCGGGCATCCGCG 498
               ||||||||||| || |||||||||||| |||||| ||||| |||| | ||||||||
Qy   221 GAGCGGACTTCCAGTGCTTCCAGTGCTTCCAGCAGCGGCCGCGGCTGGCCGCGGCACCTTCCGGG 280
          || ||||||||||||||||||||||||||||| | |||| | ||| | |||| || |||||
Db   499 GGGCCGGACTTCCAGTGTTCCAGTGCTTCCAGCAGCGGCGGCCCCGTGGGCGCTGGCGGCACCTTCCGCG 558
               |||| || ||||||||||| ||||||||||||||||||||||||||||||| || ||
Qy   281 CCTTCCTGTCCTCGCCGCTGCGAGGACCTCTACAGCATCGTGCGCGCCGCCGACCGGCACCG 340
          |||||||| |||||||||||||||| ||| |||||| | ||||||||| ||||||| |||||
Db   559 CCTTCCTGTCCTCGCCGCTGCGAGGACCTGTACAGCATCGTGCGCGCCGCCGACCGGCGCCAG 618
              ||||||||||| |||| |||||| || | || | |||||||| ||| || ||||||||
Qy   341 GGGTGCCCGTCGTCAACCTCAGGGACGAGAGTGTCTTCCCCAGCTGGGAGGCCTTATTCT 400
             |||| || || | ||| || |||| | | | ||||| ||||||||||||||| ||||
Db   619 CCGTGCCCATCGTCAACCTCAAGGACGAGAGCGTGTTCCCCAGCTGGGAGGCTCTGTTCT 678
            ||||| ||||||| ||||||||||||||||||| | ||||||| ||||||| ||| ||
Qy   401 CGGGCTCCGAGGCGCCAGCTGAAGCCCGGCCATCTCTCTTCGACGGCAGAGATG 460
             ||| |||||||| |||| |||||||| |||||||| || | |||||||||||
Db   679 CAGGCCTGAGGGTCCGTGAAGCCCGGCCACGCCATCTCTCCTTGACCGCCAAGGACG 738
             || | | ||||| |||||| |||||||| ||||||||| ||| ||||||||||||
Qy   461 TCCTGCAGCACCCGCCTGCCCCGGAAGAGCGTGTGCACGCTCCGACCCCAGCGGGC 520
             ||| | |||||| |||||||| || ||| | | |||| |||||| |||||||||
Db   739 TCCTGAGGCACCCCACCTGGCCCCAGAAGAGCGTGTGGCATGGCGTGGGCCACCGGGC 798
             ||| ||||||| | | | ||| |||||| || | |||| ||||||||||||||
Qy   521 GCCGCCTGACCGACAGCTACTGCGACGCGTGGCGACGACGGAGGCCCCGGCCGGGC 580
          || ||||||||| |||||||||| |||| ||||| ||||| |||| | ||||||||||
Db   799 GCAGGCTGACCGACAGCTACTGTGACGCGTGGCGACGGCGGAGGCTCCCTCGCCAGGGCC 858
              ||||||||||||||||||| || ||||||||||||||||||| || ||| |||| ||
Qy   581 AGGCTCGTCGCTCGTGCGGCGCCAGGCTGCTCGGAGCAGGAGGCCCGCCGACCACG 640
          ||||||| | ||| ||||||| | | ||||| |||||||||||| |||||| |||||
Db   859 AGGCCTCTCCGCTCGCTGGGGGCAGGCTCCTCGGGGCAGAGTGCCCGAGTGCGCCATCACG 918
            ||| | ||||||| | | |||||| |||||| |||| || ||| | |||||
Qy   641 CCTTCGTCGTGCTCTCGCATCGAGAAACAGCGTCATGACCTCCTTC 684
             | | | | || | | | | || || | |||||| | |||
Db   919 CCTACATCGTGCTCTGCATTGAGAAACAGCTTCATGACTGCCTCC 962

RESULT 13
ABQ76740
ID   ABQ76740 standard; DNA; 1564 BP.
XX
AC   ABQ76740;
XX
DT   03-MAR-2003 (first entry)
XX
DE   DNA encoding human Endostatin/IgG1Fc fusion construct.
XX
KW   Human; endostatin; IgG1Fc; tumour; vascular endothelial proliferation;
KW   vascular endothelial cytopoiesis inhibiting factor; inhibitor;
KW   fusion construct; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1527
FT   /*tag= a
FT   /product= "endostatin/IgG1Fc construct"
FT   /transl_except= (pos58..60,aa:Ala)
FT   /transl_except= (pos:811..816,aa:GGGGSGG)
XX
PN   CN1354186-A.
XX
PD   19-JUN-2002.
XX
PF   30-NOV-2000; 2000CN-00123347.
XX
PR   30-NOV-2000; 2000CN-00123347.
XX
PA   (LIAO-) LIAONING WEIXING BIOLOGICAL PROD INST CO.
XX
PI   Chen L, Li Z, Liu Q;
XX
DR   WPI; 2002-751441/82.
DR   P-PSDB; ABG73586.

```
XX
PT   Preparation of recombinant human vascular endothelial cytopoiesis
PT   suppressor factor with human IgG1Fc fragment molecular structure and
PT   application of its product.
XX
PS   Disclosure; Page 7-8 (Disclosure); 12pp; Chinese.
XX
CC   This invention describes a novel method for the preparation of
CC   recombinant human vascular endothelial cytopoiesis inhibiting factor with
CC   human IgG1Fc fragment molecular structure and its product application.
CC   The novel factor is derived from endostatin (using PCR to screen a human
CC   foetal kidney cell cDNA library, and human IgG1Fc. The product of the
CC   invention can specifically inhibit tumour vascular endothelial
CC   proliferation and can be used for curing several tumour types. This
CC   sequence encodes a fusion construct composed of human endostatin and
CC   human IgG1Fc, described in the disclosure of the invention
XX
SQ   Sequence 1564 BP; 346 A; 491 C; 436 G; 291 T; 0 U; 0 Other;

Query Match             51.1%;  Score 423.6;  DB 6;  Length 1564;
Best Local Similarity   77.9%;  Pred. No. 1.9e-60;
Matches  528;  Conservative    0;  Mismatches 139;  Indels  11;  Gaps   1;

Qy     137 CCCACACCCACCAGGACTTCAGCTGTGTTGCACCTGGTTGGCCCTGAACAGCCCGACAGC 196
               |||||||||||||||    |||||||| |||||||||||||||||||||
Db     260 CTCACAGCCCACCCGGACTTCCAGCCCGTGCTCCACCTGGTTGCGTCAACAGCCCCCTGT 319

Qy     197 CGGGGCGGCATGCGAGGGCATCCGGGAGGAGCGGACTTCCAGTGTTCCAGCAGGCGGGCC 256
               ||||||||||||    ||||||||||||||||||||||||||||||||||
Db     320 CAGGCGGGCATGCGGGGGCATCCGCGGGGCCGACTTCCAGTGCTTCCAGCAGGCGGGCCG 379

Qy     257 CGGGGGCTGGCCCGGCACCTTCGGGGCCTTCGTGCGTGGGCCTGCGAGGACCTCTACAGCA 316
               ||||||||||||||||   |||||||||||||||||||||||||
Db     380 TGGGGTTGGCGGGCACCTTCGGCGGCCTTCGGTCCTCGGCCTTGGCCGGACCCTGACAGCA 439

Qy     317 TCGTGCGCCGCGCGACCGCCACCGGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCT 376
               |||||||||||||||||||||||  ||||||||||||||||||||
Db     440 TCGTGCGCCGTGCCGACCGCCACCGGGCCATCGTCAACCTCAAGGACGAGCTGCTCTGT 499

Qy     377 TCCCCAGCTGGGAGGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGCCCGCA 436
               ||||||||||||||   | |||||||||||||||||||||||||
Db     500 TTCCCAGCTGGGAGGGCTCTGTTCTCAGGCTCGAGGGTCCGGTGAAGCCCGGGCACGCA 559

Qy     437 TCTTCTCTTTCGACGGCGAGATGTCCTGAGCACACCCGCCTGCCCCGGAGAGCCGTGT 496
               |||||||| ||.|||   |||||  |||||||||||||||||||||||||
Db     560 TCTTCTCTTTTGACGGCGCAAGGACGTCCTGAGGGCACCCACCTGCCCCACCTGCGAGAGCCGTGT 619

Qy     497 GGGACGGCTCCGACCCCAGCCGGCGCGCGCCTGACCGACAGCTACTGCGAGACGTGGCGGA 556
               |||||||||||||||||||||||||||||||||||||||||||||
Db     620 GGCATGGCTCGGACCCCAACCAGGGGCCAGCCTGACCGACAGCTACTGTGCAGACGTGGCGGA 679

Qy     557 CGGAGGGCCCGGCGGGCCACCCGGGCAGGCGTCGTCGTGCTGCTCTGCGGGGCAGGCTGCTGAGC 616
               |||||||||||||||   ||||||||||||||||||||||||||||
Db     680 CGGAGGCGTCCCCTCGGGCCACGGGCCCAGGCCCTCCTCGTCGGGGGCTGTCGGGGGGC 739

Qy     617 AGGAGGGCCGAGCTGCGCGACCGCCTTCTTCGTGGTTGCTCCGCATCGAGAACAGCGGTCATGA 676
               ||||||||||||||   ||||||||||||||||||||||||||||
Db     740 AGAGTGCCGCGAGCTGCGATCTCACGGGCTCCGGCTCTGTCTCTGCATTGAGAACAGCTTCATGA 799

Qy     677 CCTCCTTCTCCAAGTAGGGGGGGGGGGGGGGGCCTGGCCGGGACGCTTGCCTGCACCGTCACGTTTAAT 736
               ||||   |||||||||   ||||||||||||||||||||||
Db     800 CTGCCTCCAAGTAGTAGG------------AGGAAGCGGAGGTGGAGGGTCCGTCGACAA 848

Qy     737 GAGCATCCGCCGCCGCCGGGGGGGGGGGCCTGGCCGGGACGCTTGCCTGCACCGTCACGTTTAAT 796
               ||||||||||||   |||||||||||||||||||||||
Db     849 AACTCACACATGCCCACACGTGCCCAGCACCTGAACTCCTGGGGGGACCCGTCAGTCTTCCT 908

Qy     797 GTAATCCTCAAGAAATAA 814
               ||||||  || |
Db     909 CTTCCCCCAAAACCCAA 926
```

RESULT 14

```
AAT84485
ID   AAT84485 standard; cDNA; 4031 BP.
XX
AC   AAT84485;
XX
DT   19-NOV-1997 (first entry)
XX
DE   Mouse alpha-1 collagen cDNA.
XX
KW   Alpha-1 collagen; type XVIII collagen; cartilage degeneration; ss.
XX
OS   Mus musculus.
XX
FH   Key            Location/Qualifiers
FT   CDS            1..3867
FT                  /*tag= a
XX
PN   US563783-A.
XX
PD   01-JUL-1997.
XX
PF   01-DEC-1993;   93US-00159784.
XX
PR   01-DEC-1993;   93US-00159784.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Olsen BR, Oh SP;
XX
DR   WPI; 1997-350247/32.
DR   P-PSDB; AAW26328.
XX
PT   Nucleic acid encoding human alpha-1 collagen - for production of
PT   recombinant alpha-1 collagen, for use in the treatment of cartilage
PT   degeneration.
XX
PS   Disclosure; Fig 2; 35pp; English.
XX
CC   A cDNA clone (AAT84485) codes for a mouse novel type alpha-1 (XVIII)
CC   collagen (AAW26328) that is expressed in multiple organs, especially the
CC   liver, lung and kidney. It was isolated by screening murine 15.5- and
CC   17.5-day embryo libraries with probes based on murine type XII collagen
CC   and on human alpha-1 collagen cDNA. Isolated nucleic acids can be used to
CC   express recombinant alpha-1 collagen in transformed host cells. Claimed
CC   nucleic acid (see AAT84484) coding for human alpha-1 collagen (AAW26327)
CC   can be used to treat patients suffering from diseases associated with
CC   degradation of cartilage, and for supplementing collagen
XX
SQ   Sequence 4031 BP; 873 A; 1172 C; 1266 G; 720 T; 0 U; 0 Other;

Query Match             51.0%;  Score 423;  DB 2;  Length 4031;
Best Local Similarity   76.9%;  Pred. No. 2.2e-60;
Matches  548;  Conservative    0;  Mismatches 150;  Indels  15;  Gaps    2;

Qy       1 CCCTGGCGGGCAGAGATGACATCCTGGCCGGCCCCCGCGGCCGCTGCGCTGGACCCCAGCCCTAC 60
             |||||||||||||||||||||||||||||||||||||||||||||||
Db    3184 CCCTGGCGAGCGAGATGACATCCTGGCCCAACCCACCGCGCGTCGCGCCGCCAGCCAGCCTTAC 3243

Qy      61 CCCGGGG---CCCGGCACCAGGCGTCCTAGTGACTTCCAGTGGCTGCTGCCCCACTGGT 117
             ||||||   |||||||||||||||||||||||||||||||||||
Db    3244 CCTGGAGTTCCACATCACCACAGTTCCTATGTGACCTGCCGCGCCAGCCCGCCCCAC---- 3299

Qy     118 GGGCCCGTCCACACCCACCACCACCCACCACCCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTG 177
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db    3300 -------CCTCCCACTTGGTCTATACTCATCATCAGGACTTTCAGCAGTGTCTCCACCTGGTG 3351

Qy     178 GCCCTGAACAGCCCGCAGCCCGGCGGGCCGGGCGGGCCGGGCATGCGGGAGGCGGGACTTCCAGTGC 237
             |||||||||||||||||||||||||||||||||||||||||||||||
Db    3352 GCACTGAACACCCCCCTGTCTGGAGGCATGCGGTGTATCCGTGAGCAGATTTCCAGTGC 3411

Qy     238 TTCCAGCAGGCGGGCCGGGCGGGCCGGGCGGGCGCCCGGGCGCCCCACCTTCGTCGTCGGCCG 297
             |||||||||||||||||||||||||||||||||||||||||||||||
Db    3412 TTCCAGCAAGCCCGAGCCGGTGTGGGCGGCGTGTCGGGCACCTTCGGGCACCTTCGGGGCTTTCCGTCGTCCCTAGG 3471
```

```
Qy    298  CTGCAGGACCTCTACAGCATCGTGGCGCGCGGCGCACCGCGGCGGTGCCCGTGCGTCAAC  357
           ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   3472  CTGCAGGATCTCTATAGCATCGTGCGCGCCGTGCTGACCGGGGGTCGTGTGCCCATCGTCAAC  3531

Qy    358  CTCAGGGACGAGAGTGCTCTTCCCCAGCTGGGAGGGCCTTATTCTGGGGCTCCGAGGGCCAG  417
           || ||||||||||||| ||||||||||||  |||||||||||||||  |||| ||| ||
Db   3532  CTGAAGGACGAGAGTGCTCATCTCCCAGCTGGGGACTTCCGTGTTTCTGGGCTCCCAGGGGTCAA  3591

Qy    418  CTGAAGCCCGGGCCCGCATCTTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCC  477
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3592  GTGCAACCCGGGCCCGCATCTTTCTTTGACGGCAGAGATGTCCTGAGACACCCAGCC  3651

Qy    478  TGGCCCCGGAAGAGCGTGTGGCACGGCTCCGACCCCAGCGGGCGGCGCCTGACCGACAGC  537
           ||| ||||||||||||||||||||||||||||||||||||| ||||||||||||  |||
Db   3652  TGGCCGAGAAGAGCGTATGGCACGGCTCCGACCCCAGTGGGCGGCCGGAGGCTGATGGAGAGT  3711

Qy    538  TACTGCGAGACGTGGGGACGGAGCGCCCCGGCCGGCGCCACCGACCGGCGGTCGTCGCTGCTG  597
           |||||||||||| |||||||||||||||||||||||||||||||| |||||||||||||||
Db   3712  TACTGTGAGACATGGGCAACTGCAAACTACTGGGGTCACAGGTCAGGCCCTCCTCCCCTGCTG  3771

Qy    598  GCGGGGCAGGCTGCTGAGCAGCAGGAGGCCGCGAGCTGCCGCCCACCGCCTTGTGGTGGCTCTGC  657
           ||||||||||| |||  ||||||||| |||||||||| ||||||||||| ||  |||||||| ||
Db   3772  TCAGGGACAGGCTCCTGAGCAGCAGGAAGCTGCGGAGCTGCCACAACAGCTACACATGCTCCTGTGC  3831

Qy    658  ATCGAGAAACAGCCTCATGACCTCCTTCTCCAAGTAGGGCGCGGCGGGCCCACCG  710
           || |||||| ||||||||| |||| || ||||| |||| ||| ||  |||
Db   3832  ATTGAGAATAGCTTCATGACCTCTTTTCTCCAAATAGGCCTCTGCCAGTAGGG  3884
```

RESULT 15
AAA64013
ID AAA64013 standard; DNA; 816 BP.
XX
AC AAA64013;
XX
DT 20-DEC-2000 (first entry)
XX
DE DNA encoding a human angiogenesis inhibitor designated KED.
XX
KW Angiogenesis inhibitor; KED; kringle protein; angiogenesis; tPA;
KW tissue plasminogen activator; tumour; atherosclerosis; arthritis;
KW retinopathy; bronchial vascular congestion; inflammatory bowel disease;
KW adult respiratory distress syndrome; Castleman's disease; psoriasis;
KW hepatitis; aneurysm; renal disease; haemangioma; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   1..816
FT   /*tag= a
FT   /product= "angiogenesis inhibitor KED"
XX
PN WO200049871-A1.
XX
PD 31-AUG-2000.
XX
PF 24-FEB-2000; 2000WO-US004798.
XX
PR 24-FEB-1999; 99US-0121341P.
PR 25-FEB-1999; 99US-0122633P.
PR 18-NOV-1999; 99US-0166176P.
XX
PA (FORD-) FORD HEALTH SYSTEM HENRY.
XX
PI Dou D, Chopp M, Wang L, Mikkelsen T;
XX
DR WPI; 2000-572016/53.
DR P-PSDB; AAB08407.
XX
PT Use of kringle protein and kringle derived from plasminogen and
PT composition comprising kringle proteins for treating tumor and
PT atherosclerosis, arthritis and retinopathy.

XX
PS Disclosure; Page 151; 163pp; English.
XX
CC The present sequence encodes a human polypeptide which is a potent
CC angiogenesis inhibitor, designated KED. The KED polypeptide, kringle
CC proteins, or a kringle derived from human tissue plasminogen activator
CC (tPA) protein are used to inhibit angiogenesis. Kringle protein, KED or
CC tPA are useful for treating tumours, as well as atherosclerosis,
CC arthritis, retinopathy and other similar diseases. KED is also useful for
CC the treatment of diseases such as bronchial vascular congestion,
CC inflammatory bowel disease, adult respiratory distress syndrome,
CC Castleman's disease, psoriasis, hepatitis, aneurysm, renal disease and
CC haemangioma
XX
SQ Sequence 816 BP; 156 A; 259 C; 254 G; 147 T; 0 U; 0 Other;

Query Match   50.8%;   Score 421.2; DB 3;   Length 816;
Best Local Similarity 85.0%;   Pred. No. 4.8e-60;
Matches 471; Conservative   0; Mismatches 83; Indels   0; Gaps   0;

```
Qy    131  CCCACACCCACACCCACCCAGGACTTCCAGCTGGTGCTGCTGCACCTGGTGGCCCTGAACAGCC  190
           ||||||||||||||||| ||||||||||||||||||| ||| |||||||||| |||||||||
Db    257  CCCATATGCACAGCACCCGCGACTTCCAGCCGGTGCTCCACCTGGTTGCGCTCAACAGCC  316

Qy    191  CGCAGCCCGGGCCGGCATGCGAGGCATCCGGGGCAGGACTTCCAGTGCTTCCAGCAGGCGC  250
           ||| ||||||||| ||| ||||||| ||| ||||||||||||||| ||| |||||||||
Db    317  CCCTGTCAGGGCCGCATGCGGGGCATCCGGGGCATCCGACTTCCAGTGCTTCCAGCAGGCGGC  376

Qy    251  GGCCGCGGGGCTGGCCCGCACCTTCCGGGCCTTCGTGTCGTGGCCTGCAGGACCTCT  310
           |||||||||||| |||||||||||| |||||||||||||| ||||||||||||||||
Db    377  GGGCCGTGGGGGCTGGGGCGGCCACCTTCCGCGGCCTTCCTGCGTCCTCGCGCCTGCAGGACCCTGT  436

Qy    311  ACAGCATCGTGGCGCGCGGCGCACCGCGGCGGTGCCCGTCGTCAACCTCAGGGACGAGG  370
           |||||||||||| |||||| |||||||| ||||||||| || ||||||||||||||||
Db    437  ACAGCATCGTGCGCGCCGTGCCCATCGTCAACCTCAAGGACGAGC  496

Qy    371  TGCTCTTCCCCAGCTGGGAGGGCCTTATTCTGGGGCTCCGAGGGCCAGCTGAAGCCCGGGG  430
           |||||||||||||||||||||||||||| ||||||||||||||||| |||||||||||||
Db    497  TGCTGTTTCCCCAGCTGGGAGGGCTCGTTCTGTTTCCAGGCTCGAGGGCTGAAGCCCGGGGG  556

Qy    431  CCCGCCATCTTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGGCCCGCCCCGGAGA  490
           || || |||||||| ||||||||||||||||||||| ||||||||||||||||||||
Db    557  CACGCCATCTTCTTTCCTTTGACGGCAAGAGACGTCCTGAGGCACCACCCCACCTCGCCCCAGAGA  616

Qy    491  GCGTGTGCACGGCTCCGACCCCAGCGGGCGGCCGCCTGACCGACGAGCTACTGCGAGACGT  550
           ||||||||| || || ||| ||||||||||||||| ||||||||||||||||| |||||||
Db    617  GCGTGTGGCATGGCTCCGGACCCCAGCCCACCGGGCCAGCGGACGGGCCGACCGAGAGCTACTGTGAGACGT  676

Qy    551  GGCGGACGGAGGCCCCGGCGCACCGGACGCAGCGGCGTCGTCGCTGCTCGGCGGCAGGCTGC  610
           |||||||||||| |||||| |||| || |||||||||||||||| ||||||||||||||||
Db    677  GGGGGGACGGACGGAGGTCCCCTCGGGCACCGGCACGGGCACCGGCGGCTCCTCGCGTGCTGGGGGGGCAGGCTCC  736

Qy    611  TGGAGGAGGAGGCCGCGGAGAGCTGCCGCCCAGCCTTCGTGGTGGTCATCGAACAGCCG  670
           ||||||||||||| |||| ||||||||| |||||||||||||||| ||||||||||| ||
Db    737  TGGGGCAGAGAGTGCCGGGAGCTGCCAATCAGCGGTCTCTCGATTGAGAAACAGCT  796

Qy    671  TCATGACCTCCTTC  684
           |||||||||| |||
Db    797  TCATGACTGCCTCC  810
```

Search completed: March 29, 2004, 12:05:43
Job time : 378.576 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 10:57:20 ; Search time 83.8584 Seconds
(without alignments)
5486.086 Million cell updates/sec

Title: US-09-938-391-1
Perfect score: 829
Sequence: 1 cctcggcgggcagatgacat.........aataaaggaagccaaagag 829

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 27747546 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Query Length	DB	ID	Description
	1	496.8	59.9	3394	1	US-08-159-784-4	Sequence 4, Appli
	2	423	51.0	4031	1	US-08-159-784-1	Sequence 1, Appli
	3	418.4	50.5	563	4	US-09-231-077D-9	Sequence 9, Appli
	4	418	50.4	546	4	US-09-315-689-4	Sequence 4, Appli
	5	418	50.4	552	3	US-09-206-059-30	Sequence 30, Appl
	6	410.8	49.6	534	4	US-09-315-689-6	Sequence 6, Appli
	7	376.8	45.5	573	4	US-09-561-500-12	Sequence 12, Appl
	8	376.8	45.5	573	4	US-09-561-108-12	Sequence 12, Appl
	9	376.8	45.5	573	4	US-09-561-526-12	Sequence 12, Appl
	10	376.8	45.5	573	4	US-09-561-499-12	Sequence 12, Appl
	11	376.8	45.5	573	4	US-09-998-831-12	Sequence 12, Appl
	12	376.6	45.4	565	3	US-08-985-526-37	Sequence 37, Appl
	13	376.4	45.4	580	4	US-09-231-077D-7	Sequence 7, Appli
	14	376.2	45.4	558	3	US-09-449-293-3	Sequence 3, Appli
	15	376.2	45.4	558	4	US-09-775-325-3	Sequence 3, Appli
	16	376.2	45.4	558	4	US-09-775-174-3	Sequence 3, Appli
	17	374.4	45.2	568	4	US-09-231-077D-8	Sequence 8, Appli
	18	372	44.9	582	4	US-09-231-077D-6	Sequence 6, Appli
	19	371.2	44.8	620	4	US-09-231-077D-5	Sequence 5, Appli
c	20	65.6	7.9	741	4	US-09-252-991A-6774	Sequence 6774, Ap
	21	65.6	7.9	1557	4	US-09-252-991A-6920	Sequence 6920, Ap
	22	65.6	7.9	1590	4	US-09-252-991A-6824	Sequence 6824, Ap
	23	63.2	7.6	77536	4	US-09-410-551B-1	Sequence 1, Appli
	24	60.8	7.3	1116	4	US-09-252-991A-2102	Sequence 2102, Ap
c	25	60.8	7.3	1224	4	US-09-252-991A-1793	Sequence 1793, Ap
	26	60.8	7.3	2363	4	US-09-818-780-22	Sequence 22, Appl
	27	59.8	7.2	594	4	US-09-252-991A-12008	Sequence 12008, A
	28	59.8	7.2	2091	4	US-09-252-991A-12065	Sequence 12065, A
c	29	59.8	7.2	2181	4	US-09-252-991A-11948	Sequence 11948, A
	30	59.8	7.2	2583	4	US-09-252-991A-12036	Sequence 12036, Ap
c	31	59.4	7.2	1200	4	US-09-252-991A-1861	Sequence 1861, Ap
	32	59.4	7.2	1722	4	US-09-252-991A-2173	Sequence 2173, Ap
	33	58	7.0	15872	3	US-09-105-537-1	Sequence 1, Appli
	34	58	7.0	15872	4	US-09-091-609-1	Sequence 1, Appli
	35	58	7.0	15872	4	US-09-091-609-3	Sequence 3, Appli
c	36	56.6	6.8	12001	1	US-08-458-568A-11	Sequence 11, Appl
	37	55.8	6.7	1820	1	US-08-173-508-7	Sequence 7, Appli
c	38	55.8	6.7	1821	2	US-08-265-310-7	Sequence 7, Appli
c	39	55.8	6.7	1821	2	US-08-951-742-7	Sequence 7, Appli
c	40	55.6	6.7	4257	3	US-08-690-473-1	Sequence 1, Appli
	41	55.6	6.7	4257	3	US-09-259-821A-1	Sequence 1, Appli
	42	55.6	6.7	4257	3	US-08-843-659-1	Sequence 1, Appli
	43	55.4	6.7	12001	1	US-08-458-568A-11	Sequence 11, Appli
c	44	55.4	6.7	1203	3	US-09-086-010-1	Sequence 1, Appli
	45	55.4	6.7	2742	3	US-09-232-468A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-159-784-4

; Sequence 4, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3394
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-159-784-4

Query Match 59.9%; Score 496.8; DB 1; Length 3394;
Best Local Similarity 77.1%; Pred. No. 2e-86;
Matches 660; Conservative 0; Mismatches 157; Indels 39; Gaps 3;

```
Qy    1 CCCTGCGCGGCAGATGACATCCTGGCCGCGGCGCCCCCGCCGGCGCCGGACCCCCAGCCCTAC   60
         ||||||||||||||||||||||||||||||||| |||||| ||||||| |||||||||||||
Db 1378 CCCTGCGCGGCAGATGACATCCTGGCCGCGGCGCCCCCTGCGCCGTCGCCCAGCGCCCCAGCCCCTAC 1437
```

COMPUTER READABLE FORM:
; 　MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; 　COMPUTER: IBM PS/2 Model 50Z or 55SX
; 　OPERATING SYSTEM: MS-DOS (Version 5.0)
; 　SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; 　APPLICATION NUMBER: US/08/159,784
; 　FILING DATE: December 1, 1993
; 　CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; 　APPLICATION NUMBER:
; 　FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; 　NAME: John F. Freeman
; 　REGISTRATION NUMBER: 29,066
; 　REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; 　TELEPHONE: (617) 542-5070
; 　TELEFAX: (617) 542-8906
; 　TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; 　SEQUENCE CHARACTERISTICS:
; 　LENGTH: 4031
; 　TYPE: nucleic acid
; 　STRANDEDNESS: single
; 　TOPOLOGY: linear
US-08-159-784-1

Query Match　　51.0%;　Score 423;　DB 1;　Length 4031;
Best Local Similarity　76.9%;　Pred. No. 2.3e-72;
Matches 548;　Conservative　0;　Mismatches 150;　Indels 15;　Gaps 2;

```
Qy     1 CCCTGCGGGCAGATGACATCCTGCCCGCCCCCGCCCTGCTGACACCCCAGCCCTAC   60
         ||||||| ||||||||||||||||||||||||||  ||||||||||||||||||||||
Db  3184 CCCTGCGGAGCAGATGACATCCTGCCCAACCCACCCGGCCTCGCTGCTGACACCCCAGCCCTTAC 3243

Qy    61 CCCGGGG--CCCCGCCACCACGGCTCCTACGTGCACTTCCAGCCGCCGGCTGCCCCACTGGT  117
         |||||||  ||||||||||||||||||||  |||||||||||||||||||||||||||||---
Db  3244 CCTGGAGTTCCACATCACCACCAGTTCTATGTGCACCTGCCGCCCGCCGCCGCCCGCCCCAC--- 3299

Qy   118 GGGCCCGTCCACACCCACCCACCCACCCACCCAGGACTTCAGCTGGTGCTGCACCTGGTG  177
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3300 ------CCTCTCACTTGCTCATACTCATCATCAGGACTTTCAGCAGGACTTTCAGCATCCTGGTG 3351

Qy   178 GCCCTGAACAGCCCGCAGCCGGGCATGCGAGGGCATCCGGGGCAGCGGACTTCCAGTGC  237
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3352 GCACTGAACACCCCCCTGTCTGGAGGCATGCGTGGGGTATCCGGTGGAGCAGAGATTTCCAGTGC 3411

Qy   238 TTCCAGCAGGGGCGGCGCGCGGGGTCGGCCGCACCTTCCGGCCCTTCCTGTCGTCGCGCG  297
         ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db  3412 TTCCAGCAGGCCCGAGCCGTGGGGGCTGTCGGGGCACCTTCGGGGCACCTTCCTGTCCTCTAGG 3471

Qy   298 CTGCAGGACCTTCACAGCATGTGCGGCCGGCGCCGACCGGGGTGCCCGTGCGTGTCAAC  357
         ||||||||| ||||||||||||| ||||||||||||||||||||||||||||||||||||
Db  3472 CTGCAGGATCTCTATAGCATGGTGCGCCGTGCGACCGGGGGTCTGCCCATCGTCAAC 3531

Qy   358 CTCAGGGACGAGGAGGTGCTCTTCCCAGCTGGGAGGGGGACTCCCTGTTTTCTGGCTCCCAGGGTCAA 417
         ||||||||||||||||||||||| |||| ||| ||| |||||||||||
Db  3532 CTGAAGGACGAGGTGCTATCTCCCAGGTGGGAGAGATGTCCTGCAGCACCCCGCC 3591

Qy   418 CTGAAGCCCGGGGCCGCCATCTTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCC  477
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3592 GTGCAACCCGGGGCCCCCATCTTTCTTTCTTTGACGGCAGAGATGTCCTGAGACACCCAGCC 3651

Qy   478 TGGCCCCGGAAGAGCGTTGGCACGGCTCCAGCACCCAGCCGCCGGCCGCCTGACCGACAGC 537
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3652 TGGCCGCAGAAGAGCGTATGCCACGGCTCGGACCCCAGTGGGCGGAGGGCTGATGGAGAGT 3711

Qy   538 TACTCGACGGTGCCGACGGAGGGCCCCGCTG  597
         |||| |||||||||||||||| |||||||||
Db  3712 TACTGTGAGACATGGCGAACTGAAACTACTGGGGCTCAGATCAGCCCTCCTCCGTCTCTGCTCTGC 3771

Qy   598 GCGGGCAGCCTGCTGGAGCAGGAGGCCGCGAGCTGCCGCACGCGTCGTCTGGTGTCTCTGC 657
```

　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　RESULT 2
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　US-08-159-784-1
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　; Sequence 1, Application US/08159784
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　; Patent No. 5643783
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　; GENERAL INFORMATION:
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　; 　APPLICANT: Bjorn R. Olsen
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　; 　TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　; 　NUMBER OF SEQUENCES: 9
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　; 　CORRESPONDENCE ADDRESS:
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　; 　ADDRESSEE: Fish & Richardson
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　; 　STREET: 225 Franklin Street
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　; 　CITY: Boston
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　; 　STATE: Massachusetts
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　; 　COUNTRY: U.S.A.
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　; 　ZIP: 02110-2804

```
Db  3772 TCAGGCAGGCTCCTGGAACAGAAAAGTGCGAGCTGCCACAACAGCTACATCGTCCTGTGC 3831
          | ||||||| |||||||  |||||   |  | ||||||  |||||||  || |||||||
Qy   658 ATCGAGAGAACAGCGTCATGACCTCCTTCTCCAAGTAGGGCGCGGCCCACGG 710
          |||||||| |||||  |||||||||| ||||||||   ||||| ||||||||
Db  3832 ATTGAGAATAGCTTCATGACCTCTTTCTCCAAATAGGCCTTCTGCCAGCTAGGG 3884
```

RESULT 3
US-09-231-077D-9
; Sequence 9, Application US/09231077D
; Patent No. 6653098
; GENERAL INFORMATION:
; APPLICANT: Harding, E.I.
; APPLICANT: Violand, B.N.
; TITLE OF INVENTION: Method of producing mouse and human
; TITLE OF INVENTION: endostatin
; FILE REFERENCE: S03071-00-US
; CURRENT APPLICATION NUMBER: US/09/231,077D
; CURRENT FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/075,587
; PRIOR FILING DATE: 1998-02-23
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 563
; TYPE: DNA
; ORGANISM: human
US-09-231-077D-9

Query Match   50.5%;   Score 418.4; DB 4;   Length 563;
Best Local Similarity   84.5%;   Pred. No. 1.5e-71;
Matches   470;   Conservative   0;   Mismatches 86;   Indels   0;   Gaps   0;

```
Qy  139 CACACCCACCAGGACTTCCAGCTGGTGCTGCCACCTGGTGCCCTGAACAGCCCGCAGCCG 198
        | ||||||||| ||||||||||  ||  ||||||||  ||||||| || |||||||
Db    6 CACAGCCACCGCGACTTCCAGCCGGTGCTGCCTGCCTGGTTGCGCTCAACAGCCCCCTGTCA 65

Qy  199 GGCGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCAGCAGGCGCGGCGCCGCG 258
        |||||||||| |||||||||  ||  ||||||||  |||||||| |||||||||||||||
Db   66 GGCGGGCATGCGGGGCATCCGCGGGCGGGCATCCGGGACTTCCAGTGCTTCAGCAGGCGCGGCCGGCCGGGCCGTG 125

Qy  259 GGGCTGGCCGGCACCTTCCGGGCCTTCGTGTCGTCGGCGCTGCGAGGACCTCTACAGCATC 318
        ||||||||  ||||||||  |||||| || |||  ||  |||||||||| |||||||||||
Db  126 GGGCTGGCGGGGCACCTTCGCGCCTTCCGCGTCCTGCGAGGACCTGTACAGCATC 185

Qy  319 GTGCGCCGCGGCACCGGCACGGGGCGCCGTGCCCGTGCTCAACCTCAACTTCAAGGACGA 378
        |||||||||  ||||  ||||||||||| ||||||||||||||||||||||||||||||
Db  186 GTGCGCCGTGCCGACCGGCAGCCGTGCCCGTGCTCAACCTCAAGGACGACGAGCTGCTGTT 245

Qy  379 CCCAGCTGGAGCGCAGAGATGTCCTGCAGCACCCCGCCTGGCCCCGGAAGAGCGTGCTCTTC 438
        ||||| |||||| ||  || ||||| || |||  ||| |||||| || ||||||
Db  246 CCCAGCTGGGAGGCTCGGCTCTGTGCGGCTGGAGCGCAGAGGCAGCGCACGCATC 305

Qy  439 TTCTCTTTGACGGCAGAGATGTCCTGCAGCACCCCGAGCACCACGAGCGTGTCCTCTTC 498
        |||||||||||||||  ||||||  || ||||||||||  |||| |||||| |||||
Db  306 TTCTCCTTTGACGGCAAGGACGTCCTGAGGGACGACAGCTACTGCGAGAGCGTGTGTGG 365

Qy  499 CACGGCTCCGACCCCCAGCGGGGCGCGCCAGGGGCTGCTGCGGGGCAGCGTGGGCGGACG 558
        ||||||||||||||  ||||  || || ||||| |||  ||||||||||| |||||||||
Db  366 CATGGCTCGGACCCCCAACGGGCCAGGGCGCAGGGGCAGCGTGGGCGGACG 425

Qy  559 GAGGCCCGGCCCCAGCCCACGGGGCAGGCGTCGTCCTGCTGCGGGGCAGGCTTCTGGAGCAG 618
        ||||||  |||||||| ||||  |||||| || | ||||| ||| |||||| |||  |||||
Db  426 GAGGCTCCCGGCGCACGGGACGGCCCAGGGCTCCTGCTGCCTCGGGGGCAG 485

Qy  619 GAGGCGCGGAGCTGCGCCACGCCCTTCGTGGTGCTCGATGGAGAACAGCGTCATGACC 678
        |||||||||| |||| |||  || ||||| | ||  | || |||||| |||||||
Db  486 AGTGCGCGGAGCTGCCATCAGCGCCTCTACATCGTCCTCTGCATTGAGAACAGCGTTCATGACT 545

Qy  679 TCCTTCTCCAAGTAGG 694
        ||||  |||||| |||
Db  546 GCCTCCAAGTAATAAG 561
```

RESULT 4
US-09-315-689-4
; Sequence 4, Application US/0931S689
; Patent No. 6346510
; GENERAL INFORMATION:
; APPLICANT: Folkman, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
; FILE REFERENCE: 05213-0229
; CURRENT APPLICATION NUMBER: US/09/315,689
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-315-689-4

Query Match   50.4%;   Score 418; DB 4;   Length 546;
Best Local Similarity   85.3%;   Pred. No. 1.8e-71;
Matches   466;   Conservative   0;   Mismatches 80;   Indels   0;   Gaps   0;

```
Qy  139 CACACCCACCAGGACTTCCAGCTGGTGCTGCCACCTGGTGCCCTGAACAGCCCGCAGCCG 198
        | ||||||||| ||||||||||  ||  ||||||| || |||||  || |||||||
Db    1 CACAGCCACCGCGACTTCCAGCCGGTGCTGCCTGGTTGCGCTCAACAGCCCCCTGTCA 60

Qy  199 GGGCATGCGGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCAGCAGGCGCGGCGCCGCG 258
        ||||||||| |||||||||  ||  ||||||||  |||||||| |||||||||||||||
Db   61 GGCGGGCATGCGGGGCATCCGCGGGCGCGGCATCCGGGACTTCCAGTGCTTCAGCAGGCGCGGGCCGGCCGGGCCGTG 120

Qy  259 GGGCTGGCCGGCACCTTCCGGGCCTTCGTGTCGTCGGCGCTGCGAGGACCTCTACAGCATC 318
        ||||||||  ||||||||  |||||| || |||  |||  |||||||||| |||||||||||
Db  121 GGGCTGGCGGGGCACCTTCGCGCCTTCCGCGTCCTGCGAGGACCTGTACAGCATC 180

Qy  319 GTGCGCCGCGGCACCGGGGTGCCCGGGTGCGTGCTCAACCTCAACTTCAAGGACGA 378
        |||||||||  ||||  | ||||||  |||||||||||||||||||||||||||||
Db  181 GTGCGCCGTGCCGACGTGCCCATCGTGCCCATCCGCGAGGACGTCGGGGCATC 240

Qy  379 CCCAGCTGGAGGCCTTATTCTGCTGGTGCGGCTGCGAGGACCTCAACCTCAAGGACGTGCTCTTC 438
        ||||| |||||  | ||  |||| ||||||  ||||||||| |||||||||||  |||||
Db  241 CCCAGCTGGGAGGCTGCTGTCTTCAGGCTCTGAGGGTCGGTCCGCTGAAGCCCGGGGGCCGCATC 300

Qy  439 TTCTCTTTCGACGGCAGAGATGTCTGCAGAGATGTCCTGGAGCTGGTGCTCTTC 498
        |||||||| ||||||||  ||||| || |||| ||||| ||| ||  |||||
Db  301 TTCTCCTTTGACGGCAAGGACGTCTCGGTCTTCAGGCTCTGAGGTCGGTCCGCTGAAGCCCGGGGGCCGGGACG 360

Qy  499 CACGGCTCCGACCCCCAGCGGGGCGCCCAGGGGCTGCTGCCGGGGCTACTGCGGCGGACG 558
        ||||||||||||||  || || |||| |||  |||  ||| ||||| ||||||  |||||
Db  361 CATGGCTCGGACCCCCAACGGGCCAGCGGGGCCAGCGGGGCAGCGTGGGCGGACG 420

Qy  559 GAGGCCCGGGCGCCACCGGGCAGCGGGCGTCGTCTGGGAGCAG 618
        |||||  |||  | ||||  |||| |||| || | |||||||
Db  421 GAGGCTCCTCGGCCACGGGCCAGGGCGCCGGGGCAGGCTCGTGGGGGCAG 480

Qy  619 GAGGCGCCGAGCTGCGCCGCCACGCCCTTCGTGGTGGTGCGATCGAGACGGTCATGACC 678
        |||||||||| |||||| ||  ||| ||||| | || | | || ||||| |||||||
Db  481 AGTGCCCGGAGCTGCGCCATCACGCCCTACACGTGCTTCTGCATTGAGAACAGCTTCATGACT 540

Qy  679 TCCTTC 684
        ||| ||
Db  541 GCCTCC 546
```

RESULT 5
US-09-206-059-30
; Sequence 30, Application US/09206059
; Patent No. 6201104
; GENERAL INFORMATION:
; APPLICANT: MacDonald, Nicholas
; APPLICANT: Sim, Kim Lee

; TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and
; TITLE OF INVENTION: Proteins and Methods of Use
; FILE REFERENCE: 05213-0370
; CURRENT APPLICATION NUMBER: US/09/206,059
; CURRENT FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.0
US-09-206-059-30
; SEQ ID NO 30
;   LENGTH: 552
;   TYPE: DNA
;   ORGANISM: Homo sapiens

Query Match   50.4%;   Score 418; DB 3; Length 552;
Best Local Similarity   85.3%;   Pred. No. 1.8e-71;
Matches 466; Conservative   0; Mismatches 80; Indels   0; Gaps   0;

Qy   139 CACACCCACCAGGACTTCCAGTCGTGTCGTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 198
          ||||||||||||||||||||||| |||||||| ||||||||||||||||||||||||||||
Db   1 CACAGCCACCGCGACTTCCAGCCGTGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA 60

Qy   199 GGCGGCATGCGAGCGAGCCGGGGAGCCGGACTTCCAGTGCTTCCAGCAGGCGCGGCCGCCG 258
          ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   61 GGCGGCATCGGGGGCCATCGCGGGGCCGGACTTCCAGTGCTTCCAGGAGCGGGCGGGGTG 120

Qy   259 GGGCTGGCCGGCACCTTCGGGGCCTTCGTCGTGTCGGCGGTGCAGGACTTCTACAGCATC 318
          |||||||||||||||||||| ||||||||||| |||||||||||||||||| ||||||||
Db   121 GGGCTGGCCGGGCACCTTCGCGCCTTCGTCCTCCGCGCTGCAGGACTTGTACAGCATC 180

Qy   319 GTGCGGCCGCCGACCGGACCGGGGGTGCCGGTGCAACCTCAGGGACGGAGGTGCTCTTC 378
          ||||||||||||| |||||| || ||||||||||||| |||||||||||||||||| ||
Db   181 GTGCGGCCGGTGCCGGACCGGTGCCAGCCTGCCCATCGTCAACCTCAAGGACGAGCTGCTGCTGTTT 240

Qy   379 CCCAGCTGGGAGGCCTTATTCTCCGGGGCTTGAAGCCCGGGCCCGCATC 438
          |||||||||||||||||| ||||||||||||||||||||||||||||||
Db   241 CCCAACGGGGCAGGTGACCTGAGCGAGCTACTGTGACGGGGCCGGGCACGCATC 300

Qy   439 TTCTCTTTGACGGAGAGATGTCCTGCGACACCCCGCCTGGCCCCGGAAGAGCGTGTGG 498
          |||||||||||||||| | ||| |||||||||||||||||||| ||||||||||||||
Db   301 TTCTCCTTTGACGGCGAAGAACGTCCTGAGGCACCCCACCTGGCCCCAGAGAGAGCGTGTGG 360

Qy   499 CACGGCTCCGACCCCAGCGGGCGGCGCCGGGGCCACGGGCGTCGTCGTGCTGCTGAGCAG 558
          |||||||||||||||||||||| |||||| |||||||||||||||||||||||||||||
Db   361 CATGGCTCCGACCCCAACGGGCCCAGGTGACCGAGAGCTACTGTGTAGAGAGCGTGTGG 420

Qy   559 GAGGCCCCGGCGGGCACCGGGCCAGCAGGGC 618
          ||||||||||||||||||||||||||||||||
Db   421 GAGGCTCCCTCGGCCACGGGCCACGGGCTCCTCGCTCCTCGGGGGGCAG 480

Qy   619 GAGGCCGGGAGCTGCCCGCCACGCCTTCGTGGTGCTCTGCATCGAGAACAGCGTCATGACC 678
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 AGTGCGCGGAGCTGCCATCACGCGCTCTGCTCTGCACTCGTGCTCCTGCATTGAGAACAGCTTCATGACT 540

Qy   679 TCCTTC 684
          ||||||
Db   541 GCCTCC 546

RESULT 6
US-09-315-689-6
; Sequence 6, Application US/09315689
; Patent No. 6346510
; GENERAL INFORMATION:
;   APPLICANT: Folkman, Judah
;   APPLICANT: O'Reilly, Michael
;   TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
;   FILE REFERENCE: 05213-0229
;   CURRENT APPLICATION NUMBER: US/09/315,689
;   CURRENT FILING DATE: 1999-05-20
;   NUMBER OF SEQ ID NOS: 6
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 534

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-315-689-6

Query Match   49.6%;   Score 410.8; DB 4; Length 534;
Best Local Similarity   85.6%;   Pred. No. 4.3e-70;
Matches 457; Conservative   0; Mismatches 77; Indels   0; Gaps   0;

Qy   151 GACTTCCAGTGGTGCTGCACCTGGTGCCCTGAACAGCCCGGCGGCCATGCGA 210
          |||||||||||||||||||||||||||||| |||||||| | |||| ||||||
Db   1 GACTTCCAGCCGTGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCAGGGGCCATGCGG 60

Qy   211 GGCATCGGGGAGGAGCCGGACTTCCAGTGCTTCCAGCAGGCGCGGCCGCCGGGGCTGGCCGGC 270
          |||||||| |||||||||||||||||||| |||||||||||||||| ||||||||||||||
Db   61 GGCATCGGGGGCCGGACTTCCAGTGCTTCCAGGAGCGGGCGGGGTGGGGCTGGCCGGGC 120

Qy   271 ACCTTCGGGGCCTTCGTCGTGTCGGCGGTGCAGGACTTCTACAGCATCGTGCGGCCGCCC 330
          ||||||||||||||||||||| |||||||||||||||| ||||||||||||||||||
Db   121 ACCTTCGCGCCTTCGTCCTCCGCGCTGCAGGACTTGTACAGCATCGTGCGGCCGGTGCC 180

Qy   331 GACCGGACCGGGGTGCCGGTGCAACCTCAGGGACGAGGTGCTCTTCCCCAGCTGGGAG 390
          |||||||||||||||| || |||||||||||||||||||| ||||||||||||||||||
Db   181 GACCGGTGCCAGCCGTGCCATCGTCAACCTCAAGGACGAGCTGCTGTTTCCCAGCTGGGAG 240

Qy   391 GCCTTATTCTCGGGCTCCGAGGGCCAGTGAAGCCCGGGCCCGCATCTTCTCTTTGAC 450
          ||||||||||||||||||||| |||||||||||||||||||||||||||||||| |||
Db   241 GCTCTGTTCTCAGGGTCTGAGGGTCCGGCTGAAGCCCGGGCACGGGCCGCCCATCTTCTCCTTTGAC 300

Qy   451 GGCAGAGATGTCCTGCAGCACCCCGCCTGGCCCCGGAAGAGCGTGTGCACGGCTCCGAC 510
          |||||||| ||| ||||||||||||| |||||||| ||||||||||||||||||||||||
Db   301 GGCAAGGACGTCCTGAGGCACCCCACCTGGCCCCAGAGAGAGCGTGTGCATGGCTCCGAC 360

Qy   511 CCCAGCGGGCGCGCCGGGGCCACTACGGGGCGTCGTCGTGCTGCTGAGCAGGAGGCCGCAGC 570
          ||| |||||||||||| |||||||||||| |||||||||||||||||||||||||||||||
Db   361 CCCAACGGGCCCAGGTGACCGAGCTACTGTGAGACGGGGCCGGGCGGGCTCCTCGCTCCTCG 420

Qy   571 GCCACCGGGGCCAGGCGTCGTCGTCGTGCTGCTGAGCGAGGAGGCCGGCAGC 630
          |||||||||||||||| ||||||||||||||||||||||| | ||||| |||
Db   421 GCCACGGGCCAGGCCTCCTCGCTGCTGCGTCGTGCGGGGGGCAGAGTGCCGGCAGC 480

Qy   631 TGCCGCCACGGCCTTCGTGGTGCTCTGCATCGAGAACAGCGTCATGACCTCCTTC 684
          ||||||||||||||| |||||| |||| |||||||||||||||||||||||||| |
Db   481 TGCCATCACGCCTACATCGTGCTCCTGCATTGAGAACAGCTTCATGACTGCCTCC 534

RESULT 7
US-09-561-500-12
; Sequence 12, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
;   APPLICANT: Philip E. Thorpe
;   APPLICANT: Rolf A. Brekken
;   TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
;   FILE REFERENCE: 4001.002500
;   CURRENT APPLICATION NUMBER: US/09/561,500
;   CURRENT FILING DATE: 2000-04-28
;   PRIOR APPLICATION NUMBER: 60/131,432
;   PRIOR FILING DATE: 1999-04-28
;   NUMBER OF SEQ ID NOS: 44
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 573
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
;   OTHER INFORMATION: OLIGONUCLEOTIDE
;   NAME/KEY: CDS
;   LOCATION: (1)..(573)
US-09-561-500-12

Query Match   45.5%;   Score 376.8; DB 4; Length 573;
Best Local Similarity   79.3%;   Pred. No. 1.3e-63;

Matches 447; Conservative   0; Mismatches 117; Indels   0; Gaps   0;

Qy 126 CCACACCCACACCCACCCACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCTGGCCCTGAA 185
            || || || || || ||||||| |||| |||||||| ||||||||||| ||||||
Db   9 CCATCACCATCACCATACTCATCAGGACTTTCAGCCAGTGCTCCACCTGGTGCCACTGAA 68

Qy 186 CAGCCCGCAGCCCGGCGGGCATGCGAGGCATCCGGGAGCCGACTTCCAGTGCTTCCAGCA 245
            || ||| || || |||||||| ||||| ||| || ||| |||||||||||||||
Db   69 CACCCCCCTGTCTGGAGGCATGCGTGTATCCGTGGAGCAGATTCCAGTGCTTCCAGCA 128

Qy 246 GGCGCGGCCGCGGGGCTGGCCGGCACCTTCCGGGCCTTCGTGTCGGCTGCAGGA 305
            ||| || ||||||||||| || ||||||||||||| ||| || | |||||||
Db 129 AGCCCGAGCCGTGGGGCTGTCGGCGCACCTTCCGGGCTTTCCTGTCCTCTAGGCTGCAGGA 188

Qy 306 CCTCTACAGCATCGTGCGCGCGCGCGCCGACCGGGTGCCCGTCGTCAACCTCAGGGA 365
            |||||| |||||||||||||| |||| || ||||| |||| |||||||| ||||||
Db 189 TCTCTATAGCATCGTGCGCGCCGTGCTGACCGGGGGTCGTGTGCCCATCGTCAACCTGAAGGA 248

Qy 366 CGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCC 425
            |||||||| | ||||| |||||| ||| | | ||||||| |||| |||| |||| ||||
Db 249 CGAGGTGCTATCTCCCAGCTGGGACTCCCTGTTTTCTGGCTCCCAGGGTCAACTGCAACC 308

Qy 426 CGGGGCCCGCATCTTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCCTGGCCCCG 485
            |||||||||||||| |||| ||||||||||||||||| |||||| ||||||||| |
Db 309 CGGGGCCCGCATCTTTTCTTTTGACGGCAGAGATGTCCTGAGACACCCAGCCTGGCCGCA 368

Qy 486 GAAGAGGTGTGCACGGCTCCGACCCCGCCGCGGCCAGCGGCTGCTGCGGGCAG 545
            |||||| ||| |||||||| ||||| ||| ||| |||| || | ||||||| |
Db 369 GAAGAGCGTATGGCACGGCTCGGACCCCCAGTGGGCGGAGGCTGATGGAGAGTTACTGTGA 428

Qy 546 GACGTGGGGACGAGGGCCCCGGCGCCACCGGCCAGGCGTCGTGCTGCGGGCAG 605
            ||| || || || || || || | ||| || |||||| | |||||| |||||
Db 429 GACATGGCGAACTGAAACTACTGGGGCTACAGGTCAGGCCTCAGGGCCTGTCGTCAGGCAG 488

Qy 606 GTGCTGGAGCAGGAGGAGCCGGGAGCTGCCGCCACGGCCTTCGTGGTGCTCTTGCATCGAGAA 665
            || || |||| ||| ||| || ||||| | || || |||| || || || |||||||||
Db 489 GTCCCTGGAACAGAAAGCTGCCAGCTGCCACAACAGTACACTCGTCTGTGTGCATTGAGAA 548

Qy 666 CAGCGTCATGACCTCCTTCTCCAA 689
            || ||||||||||| ||||||||
Db 549 TAGCTTCATGACCTCTTTCTCCAA 572

RESULT 8
US-09-561-108-12
; Sequence 12, Application US/09561108
; Patent No. 6342221
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/561,108
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: OLIGONUCLEOTIDE
; NAME/KEY: CDS
; LOCATION: (1)..(573)
US-09-561-108-12

Query Match   45.5%; Score 376.8; DB 4; Length 573;
Best Local Similarity 79.3%; Pred. No. 1.3e-63;
Matches 447; Conservative   0; Mismatches 117; Indels   0; Gaps   0;

Qy 126 CCACACCCACACCCACCCACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCTGGCCCTGAA 185
            || || || || || ||||||| |||| |||||||| ||||||||||| ||||||
Db   9 CCATCACCATCACCATACTCATCAGGACTTTCAGCCAGTGCTCCACCTGGTGCCACTGAA 68

Qy 186 CAGCCCGCAGCCCGGCGGGCATGCGAGGCATCCGGGAGCCGACTTCCAGTGCTTCCAGCA 245
Db   69 CACCCCCCTGTCTGGAGGCATGCGTGTATCCGTGGAGCAGATTCCAGTGCTTCCAGCA 128

Qy 246 GGCGCGGCCGCGGGGCTGGCCGGCACCTTCCGGGCCTTCGTGTCGGCTGCAGGA 305
Db 129 AGCCCGAGCCGTGGGGCTGTCGGCGCACCTTCCGGGCTTTCCTGTCCTCTAGGCTGCAGGA 188

Qy 306 CCTCTACAGCATCGTGCGCGCGCGCGCCGACCGGGTGCCCGTCGTCAACCTCAGGGA 365
Db 189 TCTCTATAGCATCGTGCGCGCCGTGCTGACCGGGGGTCGTGTGCCCATCGTCAACCTGAAGGA 248

Qy 366 CGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCC 425
Db 249 CGAGGTGCTATCTCCCAGCTGGGACTCCCTGTTTTCTGGCTCCCAGGGTCAACTGCAACC 308

Qy 426 CGGGGCCCGCATCTTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCCTGGCCCCG 485
Db 309 CGGGGCCCGCATCTTTTCTTTTGACGGCAGAGATGTCCTGAGACACCCAGCCTGGCCGCA 368

Qy 486 GAAGAGGTGTGCACGGCTCCGACCCCGCCGCGGCCAGCGGCTGCTGCGGGCAG 545
Db 369 GAAGAGCGTATGGCACGGCTCGGACCCCCAGTGGGCGGAGGCTGATGGAGAGTTACTGTGA 428

Qy 546 GACGTGGGGACGAGGGCCCCGGCGCCACCGGCCAGGCGTCGTGCTGCGGGCAG 605
Db 429 GACATGGCGAACTGAAACTACTGGGGCTACAGGTCAGGCCTCAGGGCCTGTCGTCAGGCAG 488

Qy 606 GTGCTGGAGCAGGAGGAGCCGGGAGCTGCCGCCACGGCCTTCGTGGTGCTCTTGCATCGAGAA 665
Db 489 GTCCCTGGAACAGAAAGCTGCCAGCTGCCACAACAGTACACTCGTCTGTGTGCATTGAGAA 548

Qy 666 CAGCGTCATGACCTCCTTCTCCAA 689
Db 549 TAGCTTCATGACCTCTTTCTCCAA 572

RESULT 9
US-09-561-526-12
; Sequence 12, Application US/09561526
; Patent No. 6416758
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002586
; CURRENT APPLICATION NUMBER: US/09/561,526
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: OLIGONUCLEOTIDE
; NAME/KEY: CDS
; LOCATION: (1)..(573)
US-09-561-526-12

Query Match   45.5%; Score 376.8; DB 4; Length 573;
Best Local Similarity 79.3%; Pred. No. 1.3e-63;
Matches 447; Conservative   0; Mismatches 117; Indels   0; Gaps   0;

Qy 126 CCACACCCACACCCACCCACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAA 185

RESULT 10

```
Db      9 CCATCACCATCACCATACTCATCATCAGGACTTTCAGCCAGTGCTTCCACCTGGTGGCACTGAA     68

Qy    186 CAGCCCGCAGCCGGCGCGGCATGCGGAGGCATCCGGGAGCGGACTTCCAGTGCTTCCAGCA    245
          |||      |||||        ||   ||||||||||    |||||||||||||||||||||
Db     69 CACCCCCTGTCTGGAGGCATGCGTGGTATCCGTGGAGCAGATTTCCAGTGCTTCCAGCA    128

Qy    246 GGGCGGCGCGGCGGGGGCTGGCCGGCACCTTCCGGGGGCCTTCCTTCCTGTCGTGTCGCGGCGGA    305
Db    129 AGCCCGAGCCGTGGGGGCTGTCGGGGCACCTTCCGTGGCCTTTCCTGTCCTCTAGGCTGCAGGA    188

Qy    306 CCTCTACAGACATCGTGCGCGCCGACCGCACCGGGGTGCCGTGTCACCTGTCAACCTCAGGGA    365
Db    189 TCTCTATAGCATCGTGCGCGCGTGCTGACCGGGGTCTGTGCCCATCGTCAACCTGAAGGA    248

Qy    366 CGAGGTGCTCTTCCCAGCTGGGAGGGCCTTATTCTCGGGGCTTCCGAGGGCCAGCTGAAGCC    425
Db    249 CGAGGTGCTATCCCCAGCTGGGACTCCCCTGTTTTCTGGCTCCCAGGGTCAACTGCAACC    308

Qy    426 CGGGGCCCCGCATCTTCTTCTTCGACGGCAGAGATGTCCTGAGCACCCCGCCTGGCCCCG    485
Db    309 CGGGGGCCCGCATCTTTCTTTTGACGGCAGAGATGTCCTGAGACACCCAGCCTGGCCGCA    368

Qy    486 GAAGAGCGTGTGGCACGGCTCCGACCCCAGCGGGCGCCCTGACCGACAGCTACTGCGA    545
Db    369 GAAGAGCGTATGGCACGGCTCCGACCCCAGTGGGCAGGCGGAGGCTGATGGAGAGTTACTGTGA    428

Qy    546 GACGTGGCGGACGGAGGCCCCGGCCACCGGGCAGGCGTCGTCGCGTCTGTCGGGGCAG    605
Db    429 GACATGGCGGAACTGAAACTACTGGGGCTACAGGTCACAGTCAGGCCTCAGGCTGTCAGGCAG    488

Qy    606 GCTGCTGGAGCAGGAGGCCGCAGAGGTGCCGCCACGCCTTCGTGGTGCTCTGCATCGAAGAA    665
Db    489 GCTCCTGGAACAGAAAAGTGCGAGCTGCCACAACAGCTACATCGTCCGTGTGCATTGAGAA    548

Qy    666 CAGCGTCATGACCTCCTTCTCCAA    689
          |||||||||||||||||||||||
Db    549 TAGCTTCATGACCTCTTCTCCAA    572
```

```
Query Match             45.5%;  Score 376.8;  DB 4;  Length 573;
Best Local Similarity   79.3%;  Pred. No. 1.3e-63;
Matches   447;  Conservative  0;  Mismatches 117;  Indels  0;  Gaps  0;

Qy    126 CCACCCCACACCACCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGGCCCTGAA    185
          ||||    |||| |||| ||||||||||||||||||||||||||||
Db      9 CCATCACCATCACCATACTCATCAGGACTTTCAGCCAGTGCTTCCACCTGGTGGCACTGAA     68
```

RESULT 11

```
Qy    186 CAGCCCGCAGCCGGCGCGGCATGCGGAGGCATCCGGGAGCGGACTTCCAGTGCTTCCAGCA    245
          |||      |||||        ||   ||||||||||    |||||||||||||||||||||
Db     69 CACCCCCTGTCTGGAGGCATGCGTGGTATCCGTGGAGCAGATTTCCAGTGCTTCCAGCA    128

Qy    246 GGGCGGCGCGGCGGGGGCTGGCCGGCACCTTCCGGGGGCCTTCCTTCCTGTCGTGTCGCGGCGGA    305
Db    129 AGCCCGAGCCGTGGGGGCTGTCGGGGCACCTTCCGTGGCCTTTCCTGTCCTCTAGGCTGCAGGA    188

Qy    306 CCTCTACAGACATCGTGCGCGCCGACCGCACCGGGGTGCCGTGTCACCTGTCAACCTCAGGGA    365
Db    189 TCTCTATAGCATCGTGCGCGCGTGCTGACCGGGGTCTGTGCCCATCGTCAACCTGAAGGA    248

Qy    366 CGAGGTGCTCTTCCCAGCTGGGAGGGCCTTATTCTCGGGGCTTCCGAGGGCCAGCTGAAGCC    425
Db    249 CGAGGTGCTATCCCCAGCTGGGACTCCCCTGTTTTCTGGCTCCCAGGGTCAACTGCAACC    308

Qy    426 CGGGGCCCCGCATCTTCTTCTTCGACGGCAGAGATGTCCTGAGCACCCCGCCTGGCCCCG    485
Db    309 CGGGGGCCCGCATCTTTCTTTTGACGGCAGAGATGTCCTGAGACACCCAGCCTGGCCGCA    368

Qy    486 GAAGAGCGTGTGGCACGGCTCCGACCCCAGCGGGCGCCCTGACCGACAGCTACTGCGA    545
Db    369 GAAGAGCGTATGGCACGGCTCCGACCCCAGTGGGCAGGCGGAGGCTGATGGAGAGTTACTGTGA    428

Qy    546 GACGTGGCGGACGGAGGCCCCGGCCACCGGGCAGGCGTCGTCGCGTCTGTCGGGGCAG    605
Db    429 GACATGGCGGAACTGAAACTACTGGGGCTACAGGTCACAGTCAGGCCTCAGGCTGTCAGGCAG    488

Qy    606 GCTGCTGGAGCAGGAGGCCGCAGAGGTGCCGCCACGCCTTCGTGGTGCTCTGCATCGAAGAA    665
Db    489 GCTCCTGGAACAGAAAAGTGCGAGCTGCCACAACAGCTACATCGTCCGTGTGCATTGAGAA    548

Qy    666 CAGCGTCATGACCTCCTTCTCCAA    689
          |||||||||||||||||||||||
Db    549 TAGCTTCATGACCTCTTCTCCAA    572
```

```
Query Match             45.5%;  Score 376.8;  DB 4;  Pred. No. 1.3e-63;
Best Local Similarity   79.3%;  Conservative  0;  Mismatches 117;  Indels  0;  Gaps  0;
Matches   447;

Qy    126 CCACCCCACACCACCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGGCCCTGAA    185
          ||||    |||| |||| ||||||||||||||||||||||||||||
Db      9 CCATCACCATCACCATACTCATCAGGACTTTCAGCCAGTGCTTCCACCTGGTGGCACTGAA     68

Qy    186 CAGCCCGCAGCCGGCGCGGCATGCGGAGGCATCCGGGAGCGGACTTCCAGTGCTTCCAGCA    245
```

```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-985-526-37

Query Match     45.4%;    Score 376.6;  DB 3;  Length 565;
Best Local Similarity   79.6%;   Pred. No. 1.4e-63;
Matches  445;   Conservative   0;  Mismatches 114;  Indels   0;  Gaps    0;

Qy  135 CACCCACACCACCAGGACTTCCAGCTGGTGCGTGGCCCTGAACAGCCCGCA 194
        |||||    ||| ||||||||||||||||||    ||  ||||||||||||
Db    7 CATGCATACTCATCAGGACTTCAGCCAGTGCTCCACCTGGTGCACTGACCACCCCCCT 66

Qy  195 GCCGGCGGCATGCGAGGCATCCGGGAGCGGACTTCCAGTGCTTCCAGTGCGCGCGC 254
        |||||||||||| |||| ||||| || ||||||||||||||||||||||||||||
Db   67 GTCTGGAGGCATGCGTGTATCCGTGGCAGATTCCAGCAGTGCTTCCAGCAAGCCCGAGC 126

Qy  255 CGGGGGGCTGGCCGGCCACCTTCCGGGCCTTCCGTGCGCGGCTGCAGGACCTCTACAG 314
        |||||||||||| |||||||||||||||||||||| |||||||||||||||||||||
Db  127 CGTGGGGCGTGCGGGCCACCTTCCGGGCTTCCTGTCCCTCTAGCTGCAGGATCTCTATAG 186

Qy  315 CATCGTGCCCGCCGCCGACCGGGCGTGCCGTGTCAACCTCAGGGAGCAGGAGGTGCT 374
        ||||||||||||||||||||||||    |||||||||||||||||||||||||||||
Db  187 CATCGTGCCCGCCGCCGACCGGGGTCTTGTGCCCATCGTCAACCTGAAGGACGAGGTGCT 246

Qy  375 CTTCCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGTCGAAGCCCGGGGCCCG 434
        | ||||||||||||  ||| |||||| ||||||||| ||||||||||||||||||||||
Db  247 ATCTCCCAGCTGGGACTCCCTGTGTTTTCGGCTCCCAGGGTCAACTGCAACCCGGGGCCCG 306

Qy  435 CATCTTCTTTTGACGGCAGAGATGTCCTGCGAGCACCCCGGCCCCGGAAGAGCGT 494
        ||||||| |||||||||||||||||||||| ||||| ||  |||||||||||||||
Db  307 CATCTTTTCTTTTGACGGCAGAGATGTCCTGAGACACCCAGCGTGCCGCGCAGAAGAGCGT 366

Qy  495 GTGGCACGGCTCCGACCCCCAGCGGCGCGCCGCGACCAGCTACTGCGAGACGTGGCG 554
        |||||||||||| |||||||||| |||||||| ||||| || |||||||||| |||||
Db  367 ATGGCACGGCTCGGACCCCCAGTGGGCGGAGGCTGATGAGAGTTACTGTGAGACATGGCG 426

Qy  555 GACGGAGGCCCCGGCGCCACCGGGCCAGGCGTCGTCGTGCTGGCGGGCAGCGCTGCTGGA 614
        |||||||||  ||||||||| |||||||| ||| || |||||||| |||||  |||||||
Db  427 AACTGAAACTACTGGGGCTACAGGTCAGGCCTCTCCCGCTGTCAGGCAGGCTCCTGA 486

Qy  615 GCAGGAGGCCCGGCGAGCTGCGCCACGCCTTCGTGGTGTCGTCCATCGAGAACAGCGTCAT 674
        |||||||||| ||||||||||||||||||||||||||||||||||||||| ||||||||||
Db  487 ACAGAAAGCTGCGAGCTGCCACAACAGCTACATCGTCCTGTGCATTGATTAGCTTCAT 546

Qy  675 GACCTCCTTCTCCAAGTAG 693
        |||||| |||||||| |||
Db  547 GACCTCTTTCTCCAAATAG 565
```

```
RESULT 13
US-09-231-077D-7
; Sequence 7, Application US/09231077D
; Patent No. 6653098
; GENERAL INFORMATION:
; APPLICANT: Harding, E.I.
; APPLICANT: Violand, B.N.
; TITLE OF INVENTION: Method of producing mouse and human
; TITLE OF INVENTION: endostatin
; FILE REFERENCE: S03071-00-US
; CURRENT APPLICATION NUMBER: US/09/231,077D
; CURRENT FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: 60/075,587
; PRIOR FILING DATE: 1998-02-23
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 580
; TYPE: DNA
; ORGANISM: human
US-09-231-077D-7

Query Match     45.4%;    Score 376.4;  DB 4;  Length 580;
```

```
Db      |||    ||||||||||||||||    ||||||||||||||||||
Qy   69 CACCCCCTGTCTGGAGGCATGCTGGTATCCGTGGACAGATTTCCAGTGCTTCCAGCA 128
Db  246 GGGGCGGCCGCGGGGGCTGGCGTGGCCGGCCACCTTCCGGGCACTTCCGTGCGCGGCTGCAGGA 305
Qy  129 AGCCCGAGCCGTGGGGCGTGTCGGGCCACCTTCCGGGCTTCCTGTCCCTCTAGCTGCAGGA 188
Db  306 CCCTCACAGCATGCTGCCCGCCGCCGACACCGGGCGTGCCGTGTGCAAGCTCAGGGA 365
Qy  189 TCTCTATAGCATCGTGCCCGCCGTGCGACGTGCCATCGTCAACCTGAAGGA 248
Db  366 CGAGGTGCTCTTCCCAGCTGGGAGGCCTTATTCTCGGGCTCCAGGCTGGCTGAAGCC 425
Qy  249 CGAGGTGCTATCCCCAGCTGGGACTCCCGTGTTTTCTGGCTCCAGGGTCAACTGCAACC 308
Db  426 CGGGGCCCGCATCTTTTCTTTCGACGGCAGAGATGTCCTGCGAGCACCCGCCTGGCCCG 485
Qy  309 CGGGGCCCCGCATCTTTTCTTTTGACGGCAGAGATGTCCTGAGACACCCAGCGTGCCGCA 368
Db  486 GAAGAGCGGTGTGGCACGGCTCCGACCCCCAGCGGCGCGCCGCGACCAGCTACTGCGA 545
Qy  369 GAAGAGCGTATGGCACGGCTCGGACCCCCAGTGGGCGGAGGCTGATGAGAGTTACTGTGA 428
Db  546 GACGTGGCGGACGGAGGCCCCCGGCGCCACCCGGGCCACCGGCCGGCCACGCTGCGGCGGCAG 605
Qy  429 GACATGGGCAACTGAAACTACTGGGGCTACAGGTCAGGCCTCTCCCCCTGCTCGTGTCAGGCAG 488
Db  606 GCTGCTGGAGCAGGAGCCCGCGAGCTGCGCCACGCCTTCGTGGTGCTCTCGCTGCATCCAGAA 665
Qy  489 GCTCCTGGAACAGAAAGCTGCGAGCTGCCACAACAGCTACATCGTCCTGTGCATTGCATTGAGAA 548
Db  666 CAGCGTCATGACCTCCTTCTCCAA 689
Qy  549 TAGCTTCATGACCTCTTTCTCCAA 572
```

```
RESULT 12
US-08-985-526-37
; Sequence 37, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A.
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 565 base pairs
```

```
Qy  131  CCCACACCCACCAGGACTTCCAGCTGCTGTGCTGCACCTGGTCGTGCCCTGAACAGCC  190
              |||||| ||||||| |||| ||||  ||||| ||||||||||||| | ||||||| |
Db    5  CCCATGCTCATACTCATCAGGACTTTCAGCCAGTGCTCCACCTGGTCGCACTGAACACCC   64

Qy  191  CGGAGCCGGCGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGGCGC  250
              ||| |||||||||  |||||  |||  |||||||||||||||||||| |||
Db   65  CCCTGTCTGGAGGCATGCCGTGGTATCCGTGGACAGATTTCCAGTGCTTCCAGCAAGCCC  124

Qy  251  GCGCCGGGGGTGGCCGGCCACCTTCCGCGGCCTTCCTGTCGTGCGCGGCTGCAGGACCTCT  310
              ||||||||||||| ||||||||||   ||||||||| |  |||||||||||||||||||
Db  125  GAGCCGTGGGGCGTCGGGCACCTTCCCGGGCTTTCCTGTCCTCTAGGCTGCAGGATCTCT  184

Qy  311  ACAGCATCGTGCGCCGCGCGACCGGCACCGGGGTGCCCGTGCTCAACCTCAGGGACGAGG  370
              | |||| |||||||||||| |||||||| ||||||||| ||||||||||||||||||||
Db  185  ATAGCATCGTGCGCCGCGTGCTGACCGGGGTGTGCCCATCGTGCCCATCGTCAACCTGAGGGACGAGG  244

Qy  371  TGCTCTTCCCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGCCAGCTGAAGCCCGGGG  430
              |||||||||||||||||| ||||| ||||  | ||||||||| |||| ||||||||||
Db  245  TGCTATCTCCCAGCTGGACTCCCTGTTTTTCTGGCTCCCAGGGTCAACGTGCAACCCGGGG  304

Qy  431  CCCGCATCTCTCTTCGACGGCAGAGATGTCTGCAGCACCCGCCTGGCCCCGGAAGA  490
              |||| |||| |||| ||||||||||||||| ||||||||||||||||||||||||||
Db  305  CCCGATCTTTCTTTGACGGCAGAGATGTCCTGAGACACCCAGCCTGGCCCCGCGAAGA  364

Qy  491  GCGTGTGGCACGGCTCCGACACCCCAGCGCGGCGCCACCGGCAGCTACTGCGACACGT  550
              || | |||||| ||||||||||||||  ||||||  |||||||||| || || |||| |
Db  365  GCGTATGGCACGGCTCGGACCCCCAGTGGCGGCGGCTGGGGAGCGTGATGGAGAGTTACTGTGAGACAT  424

Qy  551  GGGAGACGGAGGAGGCCCCCGGGGCCACCGGCAGGCGTGCGCTCGTCGCTGCGGGCAGGCTGC  610
              ||   ||||||| || |||||| | ||| ||  | ||||| ||||||||||||||||||||
Db  425  GGCGAACTGAAACTACTGGGGCTACAGGTCAGGCCTCAGGCAGGCTCC  484

Qy  611  TGGAGCAGGAGGCCGCGAGCTGCCGCCACCGGCCTTCGTGCTCTCGCATCGAGAACAGCG  670
              ||||| || |||| || |||||| ||| || || ||||||||| |||||||||||||||
Db  485  TGGAACAGAGAAAGCTGCGAGCTGCCACAACAGCTACATCGTCCTGTGCATTGAGAATAGCT  544

Qy  671  TCATGACCTCTTCTCCAAGTA  692
              ||||| ||||||| ||||| |
Db  545  TCATGACCTCTTTCTCCAAATA  566
```

```
Qy  137  CCCACACCCACCAGGACTTCCAGCTGGTCTGCACCTGGTCGTGCCCTGAACAGCCCCGCAGC  196
              |||||| ||||||| |||| ||||  ||||| |||||||||||||| | |||||||  |
Db    2  CTCTACTCATCAGGACTTTCAGCCAGTGCTCCACCTGGTCGCACTGAACACCCCCCTGT   61

Qy  197  CGGGCGCATGGAGGCATCCGGGAGCGGACTTCCAGTGCTTCCAGCAGGCGCGCGCCG  256
              ||| ||||| || || |||||| |||  |||||||||||||||||| |||||||||||
Db   62  CTGGAGGCATGCCGTGGTATCCGTGGACAGATTTCCAGTGCTTCCAGCAAGCCGAGCCG  121

Qy  257  CGGGGCTGGCCGGCACCTTCCGGGCCTTCCTGTCGTGCGCGGCTGCAGGACCTCTACAGCA  316
              |||||| |||||||||||| ||  |||||| | ||| |||||||||||||||||||| |
Db  122  TGGGGCTGTGCGGGGCACCTTCCTGGGCTTTCCTGTCCTCTAGGCTGCAGGATCTCTATAGCA  181

Qy  317  TCGTGCGCCGCGCGACCGGCGTGCCCGTGCTCAACCTCAGGGACGAGGTGCTCT  376
              ||||||||||||| |||||||||||||| ||||| ||||||||||||||||  |
Db  182  TCGTGCGCCGTGCTGACCGGGTCTGCCCATCGTGCCCATCGTCAACCTGAGGGACGAGGTGCTAT  241

Qy  377  TCCCCAGCTGGGAGGCCTTTATTCTCGGGCTCCGAGGCCAGCTGAAGCCCGGGGCCCGCA  436
              ||||||||||| ||| ||| || |  |||||||||||| ||||| ||||||||||||||
Db  242  CTCCCAGCTGGACTCCCTGTTTTCTGGCTCCCAGGGTCAACGTGCAACCCGGGGCCCGCA  301

Qy  437  TCTCTCTTCGACGGCAGAGATGTCTGCAGCACCCGCCTGGCCCCGGAAGAGCGTGGT  496
              ||||| ||| |||||||||||||| ||||||||||| ||||||||||||||| |||
Db  302  TCTTTCTTTGACGGCAGAGATGTCCTGAGACACCCAGCCTGGCCCCGCGAAGAGCGTAT  361

Qy  497  GGCACGGCTCCGACACCCCAGCGCGGCGCCACCGGCTACTGCGACACGTGGCCGA  556
              ||||||||| ||| |||||||  |||||||| || ||| |||| |  || | ||
Db  362  GGCACGGCTCGGACCCCCAGTGGCGGCGGCTGGGGAGCGTGATGGAGAGTTACTGTGAGACATGGGCGAA  421

Qy  557  CGGAGGCCCCCGGGGCCACCGGCAGGCGTGCGCTCGTCGCTGCCAGGCTGCTGGAGC  616
              || |||||||| |  ||| ||  | ||||| ||||||||||||| |||||| |||||
Db  422  CTGAAACTACTGGGGCTACAGGTCAGGCCTCAGGCAGGCTCCTGGAAC  481

Qy  617  AGGAGGCCGCGAGCTGCCGCCACCGGCCTTCGTGCTCTCGCATCGAGAACAGCGTCATGA  676
              || || || || |||||| ||||| || ||||||||| |||||||||||| ||||||||
Db  482  AGAAAGCTGCGAGCTGCCACAACAGCTACATCGTCCTGTGCATTGAGAATAGCTTCATGA  541

Qy  677  CCTCCTTCTCCAAGTAG  693
              || || ||||| || ||
Db  542  CCTCTTTCTCCAAATAG  558
```

```
Qy   137  CCCACACCCACCAGGACTTCAGCTCGTGCTGCACCTGGTGGCCCTGAACAGCCCGCAGC  196
          | ||  |  | || ||  ||| ||  ||||| ||  |||| | ||   |  |    |
Db     2  CTCATACTCATCAGGACTTCAGCCGAGTGCTCCACCTGGTGCACTGAACACCCCCCTGT   61

Qy   197  CGGGCGGCATGCAGGCATCCGGGGAGGCGGACTTCCAGTGCTTCCAGGCAGGCGGCGCCG  256
          | ||| ||| ||  ||||  |||| |||||||   |||| || ||| |  |  |||||||
Db    62  CTGGAGGCATGCGTGGTATCCGTGGAGCAGATTTCCAGTGCTTCCAGCAGCCCGAGCCG   121

Qy   257  CGGGGCTGCGCGGCACCTTCCGGGCCTTCCGTCGGGCTGCGAGGACCTCTACAGCA  316
          |  ||| ||  || ||||  |||| ||| |||| ||  | ||   ||||| |  ||
Db   122  TGGGGCTGCGCGGGCACCTTCCGGGCTTTCCGTGCTCCTAGGCTGCAGGGATCTCTATAGCA   181

Qy   317  TCGTGCGCCGCGCCGACCGCACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCT  376
          |||||||||| |||| ||| |||||  |   |||| || || ||||  | ||  |||
Db   182  TCGTGCGCCGTGCTGACCGGGGGGGGTCTGTGCCCATCGTCAACCTGAAGGACGAGGTGCTAT   241

Qy   377  TCCCCAGCTGGGAGGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGCCCGCA  436
          |  ||||||| ||| |||||       | || ||||| |||| |   |  |  | |||||
Db   242  CTCCCAGCTGGGACTCCCCGTGTTTTCTGGCTCCCAGGGTCAACTGCACCCCGGGGCCCGCA   301

Qy   437  TCTTCTCTTTCGACGGCGAGAGATGTCCTGCAGCACCCCGCTGGCCCCGAAGAGCGTGT  496
          | ||| |||| || || || |  ||| |  |  |  ||||  | | | ||||  |  |
Db   302  TCTTTTCTTTTGACGGCAGAGAGATGTCCTGAGACACCCCAGCCCTGGGCCGCAGAGAGCGTAT   361

Qy   497  GGCACGGCTTCGACCCCAGCGGCGCCGCCTTGACCACGCTACTGCCAGACGTCGGCGGA  556
          ||| ||||| |||   |  ||||||||   | |  |   ||  || ||  ||| || ||
Db   362  GGCACGGCTCGGACCCCAGTGGCGCGGGAGGCTGAGGAGTTACTGTGACACATGGCGAA   421

Qy   557  CGGAGGCCCCGGCGGGGCCACCGGGCAGGGCGTCGTGCTGCTCGGCGGGCAGGCTGCTGGAGC  616
          |   || | || ||  |||  | |||  || |  || || ||  |  |||||||| |||
Db   422  CTGAAACTACTGGGGCTACAGGTCAGGCCTTCGTGGTGCTTCCCGTGTCGTGCAGGCAGGCTCCTGGAAC   481

Qy   617  AGGAGGCCGCGAGCTGCGCGCCACCGGCCTTCGTGTGGCTTCTGCATCGAGAACAGCGTCATGA  676
          |  ||  ||   ||   ||  ||  || ||  |||||||    |  | |  |  ||||| |||
Db   482  AGAAAGCTGCGAGCTGCCACAACAGCTACATCGTCCTTGCATTGAGAATAGCTTCATGA   541

Qy   677  CCTCCTTCTCCAAGTAG  693
          ||| ||  ||||| |||
Db   542  CCTCTTTTCTCCAAATAG   558
```

Search completed: March 29, 2004, 13:41:19
Job time : 89.8584 secs

```
                GenCore version 5.1.6
       Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:       March 29, 2004, 13:38:55 ; Search time 325.85 Seconds
                                          (without alignments)
                                          9471.775 Million cell updates/sec

Title:        US-09-938-391-1
Perfect score: 829
Sequence:     1 cctggcggcagatgacat........aataaaggaggagccaaagag 829

Scoring table: IDENTITY.NUC
              Gapop 10.0, Gapext 1.0

Searched:     2458946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters:        4917892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      Published Applications NA:*
           1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
           2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
           3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
           4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
           5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
           6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
           7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
           8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
           9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
          10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
          11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
          12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
          13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
          14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
          15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
          16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
          17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
          18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	829	100.0	829	10	US-09-938-391-1	Sequence 1, Appli
2	555	66.9	555	10	US-09-938-391-3	Sequence 3, Appli
3	550.4	66.4	552	14	US-10-131-24-50	Sequence 50, Appl
4	550.4	66.4	552	14	US-10-292-418-34	Sequence 34, Appl
5	503.2	60.7	4875	15	US-10-264-049-835	Sequence 835, App
6	496.8	59.9	3394	9	US-09-880-107-2178	Sequence 2178, Ap
7	493.2	59.5	4551	14	US-10-060-036-144	Sequence 144, App
8	418	50.4	546	14	US-10-042-347-4	Sequence 4, Appli
9	418	50.4	549	14	US-10-131-241-53	Sequence 53, Appli
10	418	50.4	549	14	US-10-292-418-3	Sequence 3, Appli
11	418	50.4	552	9	US-09-873-676-30	Sequence 30, Appl
12	417	50.3	551	13	US-10-080-797-2	Sequence 2, Appli
13	415.6	50.1	632	14	US-10-131-241-51	Sequence 51, Appl
14	411.6	49.7	555	12	US-10-210-172-161	Sequence 161, App
15	410.8	49.6	534	14	US-10-042-347-6	Sequence 6, Appli

16	410.8	49.6	537	14	US-10-131-241-59	Sequence 59, Appl
17	403	48.6	540	14	US-10-131-241-48	Sequence 48, Appl
18	376.8	45.5	573	9	US-09-998-831-12	Sequence 12, Appl
19	376.8	45.5	573	14	US-10-373-561-12	Sequence 12, Appl
20	376.6	45.4	565	13	US-10-036-869-37	Sequence 37, Appl
21	376.2	45.4	558	9	US-09-775-174-3	Sequence 3, Appli
22	376.2	45.4	558	9	US-09-775-325-3	Sequence 3, Appli
23	376.2	45.4	624	13	US-10-080-797-4	Sequence 4, Appli
24	376.2	45.4	624	14	US-10-422-934-70	Sequence 70, Appl
25	371.8	44.8	552	14	US-10-292-418-17	Sequence 17, Appl
26	315.2	38.0	482	12	US-10-210-172-163	Sequence 163, App
27	302.8	36.5	480	12	US-10-210-172-165	Sequence 165, App
28	263	31.7	574	14	US-10-060-036-64	Sequence 64, Appl
29	227.2	27.4	306	13	US-10-430-503-37	Sequence 37, Appl
30	171.4	20.7	5204	10	US-09-971-392-159	Sequence 159, App
31	171.4	20.7	5222	15	US-10-133-937-67	Sequence 67, Appl
32	171.4	20.7	5222	15	US-10-159-563-67	Sequence 67, Appl
33	147.6	17.8	434	14	US-10-060-036-4	Sequence 4, Appli
34	95.4	11.5	900	14	US-10-131-241-45	Sequence 45, Appl
35	76.4	9.2	18435	13	US-10-156-761-412	Sequence 412, App
36	76.4	9.2	100000	14	US-10-156-761-15103	Sequence 15103, A
c 37	76.4	9.2	9025608	14	US-10-156-761-1	Sequence 1, Appli
38	68.2	8.2	2616	15	US-10-369-493-42226	Sequence 42226, A
39	67.4	8.1	1551	12	US-10-282-122A-25857	Sequence 25857, A
40	64.6	7.8	9025608	14	US-10-156-761-1	Sequence 1, Appli
41	64.4	7.8	1728	14	US-10-156-761-4493	Sequence 4493, Ap
42	64	7.7	918	14	US-10-156-761-5655	Sequence 5655, Ap
43	64	7.7	3957	14	US-10-200-562-193	Sequence 193, App
44	64	7.7	3957	14	US-10-237-551-193	Sequence 193, App
45	64	7.7	154746	10	US-09-827-688-8	Sequence 8, Appli

ALIGNMENTS

```
RESULT 1
US-09-938-391-1
; Sequence 1, Application US/0938391
; Publication No. US20030158099A1
; GENERAL INFORMATION:
;  APPLICANT: Tong, et al.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND TREATING
;  TITLE OF INVENTION: DISORDERS INVOLVING ANGIOGENESIS
;  FILE REFERENCE: PC10790A
;  CURRENT APPLICATION NUMBER: US/09/938,391
;  CURRENT FILING DATE: 2001-08-24
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 1
;  LENGTH: 829
;  TYPE: DNA
;  ORGANISM: CANINE PRO ENDOSTATIN NUCLEOTIDE SEQUENCE
US-09-938-391-1

Query Match         100.0%;  Score 829;  DB 10;  Length 829;
Best Local Similarity   100.0%;  Pred. No. 1.7e-186;
Matches  829;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 CCCTGGCGGCAGATGACATCCTGGCCGGCCCCCCCCCCCCGCCTGCTGGACCCCCAGCCCTAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCCTGGCGGCAGATGACATCCTGGCCGGCCCCCCCCCCCCGCCTGCTGGACCCCCAGCCCTAC 60

Qy     61 CCCGGGGCCCCGCACCACGGCTTCCTACGTGCACTTCCAGCGCGGCTCGCCCCCACTGGTGGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCCGGGGCCCCGCACCACGGCTTCCTACGTGCACTTCCAGCGCGGCTCGCCCCCACTGGTGGG 120

Qy    121 CCGTTCACACCCACACCCACCACCCACCAGGACTTCCAGTTCTGGTGCTGACCTGGTGGCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCGTTCACACCCACACCCACCACCCACCAGGACTTCCAGTTCTGGTGCTGACCTGGTGGCC 180

Qy    181 CTGAACAGCCCGCAGCCCGGGCGGCATGCGAGGCATCCGGGGAGCCGGACTTCCAGTGCTTC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTGAACAGCCCGCAGCCCGGGCGGCATGCGAGGCATCCGGGGAGCCGGACTTCCAGTGCTTC 240
```

```
Qy  241 CAGCAGGCGCGGCGGGGGCTGCCGGCGGGGGCCGCCTTCCGGCCCTTCCTGTCGTCGGCGCTG     300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CAGCAGGCGCGGCGGGGGCTGCCGGCGGGGGCCGCCTTCCGGCCCTTCCTGTCGTCGGCGCTG     300

Qy  301 CAGGACCTTCTACAGCATCGTGCGCCGCGACCGCACCGGGGTGCCCGTGCCGTCGTCAACCTC    360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CAGGACCTTCTACAGCATCGTGCGCCGCGACCGCACCGGGGTGCCCGTGCCGTCGTCAACCTC    360

Qy  361 AGGGACGAGGTGCTCTTCCCCAGTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTG        420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AGGGACGAGGTGCTCTTCCCCAGTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTG        420

Qy  421 AAGCCCGGGGCCCCGCATCTTCTCTTTCGACGGCCAGAGATGTCCTGCAGCACCCCGCCTGG    480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 AAGCCCGGGGCCCCGCATCTTCTCTTTCGACGGCCAGAGATGTCCTGCAGCACCCCGCCTGG    480

Qy  481 CCCCGGAAGAGCGGTGTGCGACGGCTCCGACCCCCAGCGGGCGGCCGCCTGACCAGCAGCTAC   540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 CCCCGGAAGAGCGGTGTGCGACGGCTCCGACCCCCAGCGGGCGGCCGCCTGACCAGCAGCTAC   540

Qy  541 TGGAGACGTGGGCGGACGGAGGGCCCCGGCGGGCACCCGGGCGGTGCGTCGTGCTGCTGGGCG   600
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TGGAGACGTGGGCGGACGGAGGGCCCCGGCGGGCACCCGGGCGGTGCGTCGTGCTGCTGGGCG   600

Qy  601 GGGAACAGGCTGCTGCGGGAGCGGCAGGAGGCCGCGGGAGCTGCCGCACGCGCCTTGTGGTCTCTGCATC 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 GGGAACAGGCTGCTGCGGGAGCGGCAGGAGGCCGCGGGAGCTGCCGCACGCGCCTTGTGGTCTCTGCATC 660

Qy  661 GAGAACAGCGGTCATGACCTCCTCTTCCAAGTAGGGCCCGCGCCCCACGGACAGGCGGGG    720
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 GAGAACAGCGGTCATGACCTCCTCTTCCAAGTAGGGCCCGCGCCCCACGGACAGGCGGGG    720

Qy  721 GAGGGGGCGCCCGGAGGAGCATCCGCCCGCCCGCCCGCCCGGGGGGGCCCTTGGCCCGGGACGCTTGCCTG 780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 GAGGGGGCGCCCGGAGGAGCATCCGCCCGCCCGCCCGCCCGGGGGGGCCCTTGGCCCGGGACGCTTGCCTG 780

Qy  781 CACCGTCACGTTTAATGTAATCCTCAAGAAATAAAAGGAAGCCAAAGAG   829
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 CACCGTCACGTTTAATGTAATCCTCAAGAAATAAAAGGAAGCCAAAGAG   829
```

RESULT 2

```
Query Match          66.9%;  Score 555;  DB 10;  Length 555;
Best Local Similarity 100.0%;  Pred. No. 8.3e-122;
Matches  555;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy  139 CACACCCACCAGGACTTCCAGCTCGTGCTGCACCTGGTGCCCTGAACAGCCCGAGCCG     198
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 CACACCCACCAGGACTTCCAGCTCGTGCTGCACCTGGTGCCCTGAACAGCCCGAGCCG      60

Qy  199 GGCGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGTGCTTCCAGCGCGGCGCCCCGCG 258
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGCGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGTGCTTCCAGCGCGGCGCCCCGCG 120

Qy  259 GGGCTGCCGGCACCTTCCGGCCTTCCTGTCGTCGGCGTGCGGGCTGCAGGAGGACCTCTACAGCATC     318
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GGGCTGCCGGCACCTTCCGGCCTTCCTGTCGTCGGCGTGCGGGCTGCAGGAGGACCTCTACAGCATC     180

Qy  319 GTGCGCCGCGACCGCACCGGGGTGCCCGTGCCGGGGTGCCCGTGCCGGTGCCACCTCAGGAGGAGAGGTGCTCTTC 378
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GTGCGCCGCGACCGCACCGGGGTGCCCGTGCCGGGGTGCCCGTGCCGGTGCCACCTCAGGAGGAGAGGTGCTCTTC 240

Qy  379 CCCAGTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCCGCATC     438
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CCCAGTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCCGCATC     300

Qy  439 TTCTCTTTCGACGGCCAGAGATGTCCTGCAGCACCCCGCCTGGCCCCGGAAGAGCGGTGTGG 498
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 TTCTCTTTCGACGGCCAGAGATGTCCTGCAGCACCCCGCCTGGCCCCGGAAGAGCGGTGTGG 360

Qy  499 CACGGCTCCGACCCCCAGCGGGCGCCGCCTACTGCGCAGCAGCTACTGGAGACGTGGGCGGACG 558
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 CACGGCTCCGACCCCCAGCGGGCGCCGCCTACTGCGCAGCAGCTACTGGAGACGTGGGCGGACG 420

Qy  559 GAGGGCCCCGGCGGGCACCCGGGCAGGCGGTGCGTCGTGCTGCTGGGCGGGGAGCAG    618
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GAGGGCCCCGGCGGGCACCCGGGCAGGCGGTGCGTCGTGCTGCTGGGCGGGGAGCAG    480

Qy  619 GGGCCGCGAGCTGCCGCACGCGCCTTCGTTGGTGCTGCATCGAGAACAGCGGTCATGACC   678
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 GGGCCGCGAGCTGCCGCACGCGCCTTCGTTGGTGCTGCATCGAGAACAGCGGTCATGACC   540

Qy  679 TCCTTCTCCAAGTAG   693
        |||||||||||||||
Db  541 TCCTTCTCCAAGTAG   555
```

RESULT 3

```
Query Match          66.4%;  Score 550.4;  DB 14;  Length 552;
Best Local Similarity 99.8%;  Pred. No. 1e-120;
Matches  551;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy  139 CACACCCACCAGGACTTCCAGCTCGTGCTGCACCTGGTGCCCTGAACAGCCCGAGCCG     198
        ||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db    1 CACACCCACCAGGACTTCCAGCTCCAGCCCGTGCTGCACCTGGTGCCCTGAACAGCCCGAGCCG 60

Qy  199 GGCGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGTGCTTCCAGTGCTTCCAGCGCGGCGCCCCGCG 258
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGCGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGTGCTTCCAGTGCTTCCAGCGCGGCGCCCCGCG 120

Qy  259 GGGCTGCCGGCACCTTCCGGCCTTCCTGTCGTCGGCGTGCGGGCTGCAGGAGGACCTCTACAGCATC     318
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  121 GGGCTGGCGGCGGCACCTTCCGGGCCTTCCGGTGCCCGGGTGCCGTCGTCAACCTCAGCGACGAGGTGCTCTTC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  121 GGGCTGGCGGCGGCACCTTCCGGGCCTTCCGGTGCCCGGGTGCCGTCGTCAACCTCAGCGACGAGGTGCTCTTC

Db  319 GTGCGGCGGCGGCGACCGCACCGGGGGTGCCCGGGGTGCCCGTCGTCAACCTCAGCGACGAGGTGCTCTTC 378
        |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Qy  181 GTGCGGCGGCGGCGACCGCACCGGGGGTGCCCGGGGTGCCCGTCGTCAACCTCAGCGACGAGGTGCTCTTC 240

Db  379 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGCCCGCATC 438
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  241 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGCCCGCATC 300

Db  439 TTCTCTTTCGACGGCGCAGAGATGTCCTGCGACGACCCCGCCTGGCCCCGGGAGAGCGTGTGG 498
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  301 TTCTCTTTCGACGGCGCAGAGATGTCCTGCGACGACCCCGCCTGGCCCCGGGAGAGCGTGTGG 360

Db  499 CACGGCTCCGACCCCAGCGGCGCGCCGCTACTGCGACAGCGTGGCGGACG 558
        |||||||||||||||||||||||||||||||||||||||||||||||||
Qy  361 CACGGCTCCGACCCCAGCGGCGCGCCGCTACTGCGACAGCGTGGCGGACG 420

Db  559 GAGGCCCCGGCGGCCACCGGGGCCAGGGCTGCTCGGGCGGCAGCTGCTGGAGCAG 618
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  421 GAGGCCCCGGCGGCCACCGGGGCCAGGGCTGCTCGGGCGGCAGCTGCTGGAGCAG 480

Db  619 GAGGCCGCGAGCTGCCGCCACGCCTTCGTGGTGTCATCGAGAACAGCGTCATGACC 678
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  481 GAGGCCGCGAGCTGCCGCCACGCCTTCGTGGTGTCATCGAGAACAGCGTCATGACC 540

Db  679 TCCTTCTCCAAG 690
        ||||||||||||
Qy  541 TCCTTCTCCAAG 552
```

RESULT 4
US-10-292-418-34
; Sequence 34, Application US/10292418
; Publication No. US2003013936SA1
; GENERAL INFORMATION:
; APPLICANT: Lo, Kin-Ming
; APPLICANT: Li, Yue
; APPLICANT: Gillies, Stephen D
; TITLE OF INVENTION: Expression and Export of Angiogenesis Inhibitors as
; TITLE OF INVENTION: Immunofusins
; FILE REFERENCE: LEX-006C1
; CURRENT APPLICATION NUMBER: US/10/292,418
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: 09/383,315
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: US 60/097,883
; PRIOR FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Canis familiaris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(552)
; OTHER INFORMATION: Endostatin
US-10-292-418-34

Query Match   66.4%; Score 550.4; DB 14; Length 552;
Best Local Similarity 99.8%; Pred. No. 1e-120;
Matches 551; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

```
Qy  139 CACACCCACCAGGACTTCAGCTGGTGTGCACCTGGTGGCCCTGAACAGCCCGACGCCG 198
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 CACACCCACCAGGACTTCAGCCGTGCTTCCAGCCCTGGTGGGCCCTGAACAGCCCGACGCCG 60

Qy  199 GGGGCGGCATGCGGAGGGCATCGGGGAGGCGCTTCCAGTGCTTCCAGCAGGGCGCGCCGCG 258
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GCGGCGGCATGCGGAGGGCATCGGGGAGGCGCTTCCAGTGCTTCCAGCAGGGCGCGCCGCG 120

Qy  259 GGGCTGGCGCCGGCACCTTCCGGGCCTTCCGGTGCGCGGGACCTTCTTACAGCATC 318
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GGGCTGGCGCCGGCACCTTCCTGTCGTGCGGGTGCGCAGGACCTTCTTACAGCATC 180
```

RESULT 5
US-10-264-049-835
; Sequence 835, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 835
; LENGTH: 4875
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-264-049-835

Query Match   60.7%; Score 503.2; DB 15; Length 4875;
Best Local Similarity 77.6%; Pred. No. 1.3e-109;
Matches 664; Conservative 0; Mismatches 153; Indels 39; Gaps 3;

```
Qy    1 CCCTGGCGGGCAGATGACATCCTGGCCGCGCCCCGCCCCCTGCGCCCCTCGCTGACCCCCAGCCCTAC 60
        |||||||||||||||||||||||||||||||| || |||||||||||||||||||||||||||||||
Db 2386 CCCTGGCGGGCAGATGACATCCTGGCCGCGCCCCGCCCCCTGCGCCCCTCGCTGACCCCCAGCCCTAC 2445

Qy   61 CCCGGGGCCCCGCGCACCAGGCGTCCTACGTGCACTTCCAGCCGGCTCCCCACTGGTGGG 120
        |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db 2446 CCGGAGCCCCGCGCACCAGGCGTCCTACGTGCACTTCCAGCCGGCGCGGACGCCACCACAAG--- 2501

Qy  121 CCCGTCCACACCCACACCAGGCTTCCAGTGCTTCCAGCAGGGCGGCGCCGCG 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2502 -------CCCACCCGCCACCAGCCACCGGCATCCGGGGAGGCATCCGGGGAGTGCTTCCACCTGGTTGCG 2553

Qy  181 CTGAACAGCCCGCAGCCCGGGGCGGCGGCATGCGGAGGCATCCGGGGAGCTTCCAGTGCTTC 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GCGGCGGCATGCGGAGGGCATCGGGGAGGCGCTTCCAGTGCTTCCACAGTGCTTC 120

Qy  259 GGGCTGGCGCCGGCACCTTCCGGGCCTTCCGGTGCGCGGGACCTTCTTACAGCATC 318
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2554 CTCAACAGCCCCCGTGTCAGCCGGCATCCGGGGCATCCGGGGGCCGACTTCCAGTGCTTC 2613
```

```
Query Match       59.9%;   Score 496.8;   DB 9;   Length 3394;
Best Local Similarity   77.1%;   Pred. No. 4.2e-108;
Matches  660;   Conservative   0;   Mismatches 157;   Indels  39;   Gaps  3;

Qy      1 CCCTGCGGGCAGATGACATCCTGGCCGGCCCCCCGGCCCTGCTGGACCCCAGCCCTAC   60
            |||||  |||||||||||||  ||||||||| |||| || |  |||||||||||||||
Db   1378 CCCTGGGGGGCAGATGACATCCTGGCCAGCCCCCCTGGGCTGCCCGAGCCCAGCCCTAC 1437

Qy     61 CCCGGGGCCCCGGCCACCACCAGGACTTCCAGGTGGTGCCGCCGGCTGCCCCACTGGTCGG  120
            | ||||||| | |||||  | ||||| ||||||||||||||||||||||||||||||
Db   1438 CCCGGAGGCCCGGCCACCACCAGCTGCACTGCGGGCCGGCACGACCCACAAG----     1493

Qy    121 CCCGTCCACACCCACACCACCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCC       180
            |||||||||||||||     |||||||||||||||||||||||||||||||||||
Db   1494 -------CCCACCCGCCCACAGCCACCGGCACCTTCCAGCCGGTGCTCCACCTGGTTGCG 1545

Qy    181 CTGAACAGCCCGCAGCCGGGCATGCAGCCGGGCATCCGGGAGGAGGGAGCTTCCAGTGCTTC  240
            ||| ||||||||||  |||||||  |||||||| ||||||||||||||||||||||||
Db   1546 CTCAACAGCCCCCTGTCAGGCGGCATGCGGGGCATCCGCGGGCGCACTTCCAGTGCTTC 1605

Qy    241 CAGCAGCGCGCGCCCCGCGGCGTGGCCGGCACCTTCGGCGTGCTCGTCGGCTG         300
            ||||||||| ||||||  |||||||||||||||||||||||| || |||||||||
Db   1606 CAGCAGGCGGCGGCCGGCGTGGCGGTGGCTCGGCACCTTCCGGTCCTCGTCGGCCTG 1665

Qy    301 CAGGACCTCTACAGCATCGTGCGCCGCGCGACCGGCGGTGCCCGTCGTCAACCTC        360
            |||| |||| |||| |||||||||||||||||||| |||| |||||||||||||||
Db   1666 CAGGACCTGTACACGATCGTGCGCCGCGACCGCGCGCCCATCGTCAACCTC 1725

Qy    361 AGGGACGAGGTGCTCTCCCAGCTGGGAGGCCTTATTCTGGGGCTCCGAGGGCCAGCTG    420
            ||||| ||||||||||||| ||| |||||||||| | ||||||||| |||||||||||
Db   1726 AAGGACGAGCTGCTGTTTCCCAGCTGGAGGCTCGTGTTCTCAGGCTGTGAGGGTCCGCTG 1785

Qy    421 AAGCCCGGGCCCGCATCTTCTTTCGACGGGCAGAGATGTCCTGCAGCACCCGCCTGG    480
            ||||||||| |||| ||| ||||  |||| |||| ||||| |||||||||| |||||
Db   1786 AAGCCCGGGGCACGACATCTTCTTTGACGGCCAAGGACAGTCCGAGGCACCCCACCTGG 1845

Qy    481 CCCCGGAAGAGCGTGTGGCACGGCTTCCGACCCAGCCGGCCGCGCCCGCCTGACCGACAGCTAC  540
            ||||| ||||| ||||||||||| | |||||||||||||||||||||| |||||||||||
Db   1846 CCCCAAAGAGCGTGTGGCATGGCTCGGACCCCAAGGGCGCAGGCTGACCGAGAGCTAC 1905

Qy    541 TGCGAGACGTGGCGACGGAGCATCCGCCGCCTTCGTGCTGCCTGCG         600
            ||||||||||| |||| |||||  |||| ||||||||||||||||
Db   1906 TGTGAGACGTGGCGGACGGAGCGTCCTCGGCCACAGGGCTCCTGCTCCGG 1965

Qy    601 GGCAGCGTGCTGGAGCAGGAGGCCCGGAGCTGCCGGAGCTGCCGCCACACGCGTCCGCCATC  660
            |||||||||||||||||||||||||||||||| |||||  ||||  ||||||||||||
Db   1966 GGCAGGTGCTGGACGAGGAGGCCCGGAGCTGCGAGCTGCCATCACGTCTGTCTGCATT 2025

Qy    661 GAGAACAGCGTCATGACCTCCTTCCAAGTAGGGCCGCCACCGCGGGGAGG         720
            ||||||||| ||||||||||||||||| | |||||||||||||||||||
Db   2026 GAGAACAGTTCATGACTGCCTCCAAGTAGCCACCGCGCTGGATGCAGATGGCCGGAGAGG 2085

Qy    721 GAGGGGGCGCCCGAGGAGCATCCGCGCCCGGGG--------------       756
            |||||||||||||||||||||||||||||||
Db   2086 ACCGGCGGCTCGGAGGAAGCCCCACCGTGGGCAGGGAGCAGCCCGGCCCCTGGCCC 2145

Qy    757 --GGGCCTGGCGGGGACGCTTGCCTGCACCG-TCACGTTAATGTAATCCTCAAGAAATA  813
                |||||||| | ||| |||||  ||| | |||||||||||||||||||||||||
Db   2146 CAGGACCTGGCTGGCTGCCATACTTTCCTGTATAGTTCACGTTCATGTAATCCTCAAGAAATA 2205

Qy    814 AAAGGAAGCCAAAGAG 829
            ||||||||||||||||
Db   2206 AAAGGAGGCCAAAGAG 2221
```

```
Qy    241 CAGGAGGCGGCGGCCCGCGGGGGGCTGGCCGGCACCTTCGGCGGCCTTCGGTGTCGGGGCTG  300
            ||||||| |||||||||  |||| |||||||||||||||||  ||| ||||||||||||
Db   2614 CAGCAGGCGGCGGCCGCCGTGGGGCTGGCGCCGCCCTTCGGGGCACCTTCCCGCCGCCTG 2673

Qy    301 CAGGACCTCTACAGCATCGTGCGCCGCGACCGGCGGTGCCCGTCGTCGTCAACCTC  360
            ||||||||| ||| ||||||||||||| ||||||| | |||| ||||||||||||
Db   2674 CAGGACCTGTACAGCATCGTGTACGCCGTGCCCATCGTCGACCAGCAGCTGCAACCTC 2733

Qy    361 AGGGACGAGGTGCTCTTCCCAGCTGGGAGGCCTTATTCTCTGGGGCTCCGAGGGCCAGCTG  420
            ||||| ||||||| || |||||| |||||||||| |  |||||||| |||||||||||||
Db   2734 AAGGACGAGCTGCTGTTTCCCAGCTGGGAGGCTCGTTCTCAGGCTCGTGAGGGTCCGCTG 2793

Qy    421 AAGCCCGGGGCCCGCATCTTCTTTCGACGGGCAGAGATGTCCTGCAGCAGCACCCCGCCTGG  480
            ||||||||| ||||||| |||||  ||||| |||| |||| ||||||| |||||| |||||
Db   2794 AAGCCCGGGCACCGCACCGCCATCTTCTTTGACGGCCAAGGACAGTCCGAGGCACCACCACCTGG 2853

Qy    481 CCCCGGAAGAGCGTGTGGCACGGCTTCCGACCCAGCCGGCCGCGCCCTGACCGACAGCTAC 540
            ||||| ||||| ||||||||||||| | |||||||| ||||||||||  |||||||||||
Db   2854 CCCCAGAAGAGCGTGTGGCATGGCTCGGACCCCAAGGGCGCAGGCGGCTGACCGAGAGCTAC 2913

Qy    541 TGCGAGACGTGGCGGACGGAGCAGGAGGCCCGGAGCCGTCGTCCTGCTGGCG  600
            |||||||||| ||||||||||  |||||||  ||| | ||| |||||| |||
Db   2914 TGTGAGACGTGGCGGACGGAGCGTCCTCGGCCCACAGGCCTCGTCCTGTTGGGG 2973

Qy    601 GGCAGCGTGCTGGAGCAGGAGGCCGCCGAGCTGCCCGCCACATGGCGTGCGGTCGTCTGCATC  660
            |||||||||||||||||||||||  ||||||||||  |||||||||||| ||||  |||
Db   2974 GGCAGGCTCCTGGGGACGAGGAGGCTGCTGGAGCTGCCGATCACGTCTGTCTGCATT 3033

Qy    661 GAGAACAGCTTCATGACTCCTTCTCCAAGTAGCCACCACCGCGCGCGGGG         720
            |||||||||||||||| || |||| |||||||| || |||||||| |||
Db   3034 GAGAACAGTTCATGACTGCCTCCTCCAAGTAGTGCCGCTGGATGCGGTGGCCGGAGAGG 3093

Qy    721 GAGGGGGGCGCCCGAGGAGCATCCGCGCCCGGGG---------         756
            |||||||||||||||||||||||||||||||||
Db   3094 ACCGGCGGCTCGGAGGAAGCCCCACCGTGGGCAGGGAGCAGCGCGGCCCGGCCC 3153

Qy    757 --GGGCCTGGCGGGGACGCTTGCCTGCACCG-TCACGTTAATGTAATCCTCAAGAAATA 813
                |||||||| | ||| |||||  |||| | ||||||||||||||||||||||||
Db   3154 CAGGACCTGGCTGCCTGCCATACTTTCCTGTATAGTTCACGTTCATGTAATCCTCAAGAAATA 3213

Qy    814 AAAGGAAGCCAAAGAG 829
            ||||||||||||||||
Db   3214 AAAGGAAGCCAAAGAG 3229
```

Query Match   59.5%;   Score 493.2;   DB 14;   Length 4551;
Best Local Similarity   83.9%;   Pred. No. 2.9e-107;
Matches   574;   Conservative   0;   Mismatches   98;   Indels   12;   Gaps   1;

```
Qy     1 CCCTGGCGGCGAGTGACATCGGCGGCCCCCCGCGCCTGCTGACACCCCAGCCCTAC   60
           ||||| ||||||| ||||||  |||| || ||| ||||||  ||  |||||||||||
Db  3874 CCCTGCCGGCGAGATGACATCCGGCCAGCCCCTGCGCCTGCCCGAGCCCCAGCCCTAC 3933

Qy    61 CCCGGGGGCCCCGACCACGGCTCCTACGTGCACTTCAGCCGGCTCGCCCCACTGTGGG  120
           ||||  | ||||||||||  || |||||||||||    ||  |  || |||||| |||
Db  3934 CCCGGAGCCCCGACCACCAGCTGCTACGTGCACTTGCCGCGCCGACCCCACAAG---- 3989

Qy   121 CCCGTCCACCCACCCACACCCACACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCC  180
                  ||||||| ||| ||| ||| |||||||||||| |||||||| |||||| ||
Db  3990 -------CCCACCCGCCACAGCCACCCGGACTTCCAGCCGGTGCTGCCACCTGGTTGCG 4041

Qy   181 CTGAACAGCCCGCAGCCCGGCATGCGAGGCATCCGGGAGGAGGCCGACTTCCAGTTGCTTC 240
           || ||||||| |  |||  |||||| ||  |||| |||| | |||||||||||||| ||||
Db  4042 CTCAACAGCCCCCTGTCAGCGGGCATGCGGGGCATCCGGGGCCGGGCCGACTTCCAGTTCCTTC 4101

Qy   241 CAGGAGGCGCCCGACCACGGCTCCTACGTGCACTTCCTGTCGTCGGCGTG  300
           ||||| || ||| || || || ||||||||||||||||||| || |||||
Db  4102 CAGCAGGCGGCGGCGGGCGTGGGGCGTGGGGGCGTGGGGCCGTCCTGTCGCCGTG 4161

Qy   301 CAGGACCTCTACAGCATCGTGCGCCGGCGCCGGACCGGGTGCCGGTCGTCAACCTC  360
           ||| ||||| || ||||||||| ||| ||||  ||| || |||| || || ||||
Db  4162 CAGGACCTGTACAGCATCGTGCGGCGGCCGTCCGCAGCCGTGCCCATCGTCAACCTC 4221

Qy   361 AGGGACGAGGTGCTCTTCCCCAGCTGGAGGGCCTTATTCTCGGGCTCCGAGGGCCAGGTG  420
           ||| ||||||||||| || || ||||  || ||||||| | |||| | |||||| | ||
Db  4222 AAGGACGAGCTGCTGTTTTTCCCAGCTGGAGGCTCTGTTTCTCAGGCTCTGAGGGTCCGCTG 4281

Qy   421 AAGCCCGGGGCCCCATCTTCTCTTTTCGACGGCAGATGTCTGCGAGCACCCCGCCTGG  480
           ||||| |||| || || ||||| ||||||| | || || || |||||||||  || ||
Db  4282 AAGCCCGGGGCACGAGCATCTTCTCTTTTGACGGCGTCCTGGAAGGCAAGGACTCCACCTGG 4341

Qy   481 CCCCGGAAGAGCCGTGTGGCACGGCTCCGACCCCAGCCGGCGGCGCCCTTGACCGACAGCTAC 540
           ||||| ||||||||| ||||| || ||||||||||| || || ||| ||||||||||||||||
Db  4342 CCCAGAAGAGCCGTGTGCATGGCGTCGGACCCCAACGGGGCCGAGCCGGCTGACCGAGAGCTAC 4401

Qy   541 TGCGAGACGTGGCGGACGACGGAGCCCCGGCGCCCGCAGGCGTCGTCGCTGCGGCGG  600
           | |||||||| ||||| ||||  |||  |||| ||||||  ||||||||| |  |||
Db  4402 TGTGAGACGTGGCGGACCGGACGGAGGCTCCCTCGGCGTCCTCGCACCCCGCTGTGTGGGG 4461

Qy   601 GGCAGGCTGCTGAGACGAGCAGGAGGCCGGAGCTGCGCCACGCGTTCGTGGTGCTCTGCATC  660
           |||||||||||| | || || ||||| || ||||||||||||||||||||||||||||||||
Db  4462 GGCAGGCTGCTGGAGTGCGGCGAGTGCGCCGGAGCTGCGCCACGCGTTCGTGGTGCTCTGCATT 4521

Qy   661 GAGAACAGCTCATGACCTCCTTC  684
           |||||||||||||||| |||||
Db  4522 GAGAACAGCTTCATGACTGGCCTCC 4545
```

Query Match   50.4%;   Score 418;   DB 14;   Length 546;
Best Local Similarity   85.3%;   Pred. No. 1.8e-89;
Matches   466;   Conservative   0;   Mismatches   80;   Indels   0;   Gaps   0;

```
Qy   139 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCCCTGAACAGCCCGCAGCCG  198
           |||| |||| |||||||||||| |||||||||| ||  | ||| |||||||| |||| |
Db     1 CACAGCCACCCGGACTTCCAGCCGGTGCTGCCACCTGGTTGCGCTCAACAGCCCCCTGTCA   60

Qy   199 GGCGGCATGCGAGGCATCCGGGAGGCCGGCCTTCCAGTGCTTCCAGCAGGCGCGCGCGCG  258
           |||| |||||| ||||| || ||| ||||| ||||||| || ||||||||| || || ||
Db    61 GGCGGCATGCGGGGCATCCGGGAGGCCGGCCGACTTCCAGTTCCTTCCAGCAGGCGGCGGCGCG  120

Qy   259 GGGCTGGCCGGCACCTTCGTGTCGTCCGGGCTGCGGCGGGCTGCCGCTGGTGCGGGGGGCGGTG 318
           |||||||| ||||| ||| || || |||| ||| | |||  ||| |||| |||||| | |||||
Db   121 GGGCTGGCCGGCACCTTCCGTCCGTCCGCCTGCTCAACCTCAAGGACGAGCTGCTGTTTTACAGCATC 180

Qy   319 GTGCGGCCGCGCGACCGCACCGGGCGCGCCGGCAGCCGTGCCCATCGTCAACCTCAAGGACGAGCTGCTCTC 378
           ||||| ||| || ||||| ||||| |||||| |||| ||||||||||| || ||||||||||| |||||| |
Db   181 GTGCGGCCGTGCCACCGCACCGGGCGCGCCGGCAGCCGTGCCCATCGTCAACCTCAAGGACGAGCTGCTGTTT 240

Qy   379 CCCAGCTGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCCGCATC  438
           |||||||||||| |||| | | |||| | |||||| |||||||||||||| ||||| |||
Db   241 CCCAGCTGGAGGCTCTGTTTCTCAGGCTCTGAGGGTCCGCTGAAGCCCGGGGCACGCGCATC 300

Qy   439 TTCTTTTGCACGGCAGAGATGTCCTGCGAGCACCCCCGCCTGGCCCCGGCGCGGTGTGG  498
           ||||| |||| | ||||||||| || ||||||||| ||||  || ||| ||| ||||||
Db   301 TTCTCTTTTGACGGCGTCCTGGAAGGCAAGGACTCCACCTGGCCCCAGAAGAGCCGTGTGG 360

Qy   499 CACGGCTCCGACCCCAGCGGGCGGCGCCTGCTGGAGTGCGGCGACGACGAGTGCGCGGACG 558
           |||||||||||||||| | || || ||||| ||||||||||| || ||| ||||||||||
Db   361 CATGGCTCGGACCCCAACGGGGCCGAGCCGGCTGACCGAGAGCTACTGTGAGACGTGGCGGACG 420

Qy   559 GAGGCCCCGCGCGGCCACCGGCAGGCGTTCGTGCTGCTCGGGGCGCAGGCTGTGGACGAG 618
           ||||| || || || ||||||||| || || || || ||  | |  |||||||||||||
Db   421 GAGGCTCCCTCGGCGTCCTCGCACCCCGCTGTGTGGGGGGCCAGGCTGCTGGAGTGCGGCGAG 480

Qy   619 GAGGCCGCGAGCTGCGCCACGCGTTCGTGGTGCTGCTCTGCATCACGCCTACATCGTGTCTCTGATT 678
           |||| | ||||| |||||||||||||||||||| | | |||| || |||  ||| |||| ||||| |
Db   481 AGTGCCGCGAGCTGCGCCACGCGTTCGTGGTGCTCTGCATTGAGAACAGCTTCATGACT 540

Qy   679 TCCTTC  684
           ||||||
Db   541 GCCTCC  546
```

RESULT 8
US-10-042-347-4
; Sequence 4, Application US/10042347
; Publication No. US2003011437OA1

RESULT 9
US-10-131-241-53
; Sequence 53, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer
; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-131-241-53

Query Match   50.4%;   Score 418; DB 14;   Length 549;
Best Local Similarity 85.3%;   Pred. No. 1.8e-89;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 139 CACACCCACCAGGACTTCAGCTGGTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 198
         ||||||||||||||||||||||||||||| || | |||||||| |||||||||||
Db   1 CACAGCCACCGCGACTTCAGCGTGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA 60

Qy 199 GGGCGGCATGCGAGGCATCCGGGAGCGGACTTCCAGTGCTTCCAGCAGGCGCGCGCGG 258
         |||||||||||||| |||||||||||| ||||||||||| | |||||||||||||||
Db   61 GGCGGCATGCGAGGCATCCGGGGGCGGACTTCCAGTGCTTCCAGCAGGCGCGGGCCGTG 120

Qy 259 GGGCTGGCCGGCACCTTCCGGGGCCTTCGTGTCGCGGCTGCAGGACCTTCTACAGCATC 318
         ||||||||||||||||||||||||||||||||||||||| ||||||||| |||||||||
Db 121 GGGCTGGCCGGCACCTTCCCGCGCCTTCGTGTCCGCGGCTGCAGGACCTGTACAGCATC 180

Qy 319 GTGCGCGCGGCGACCGCACCGGGGTGCCCGTCGTGTCAACCTCAGGGACGAGGTGCTCTTC 378
         ||||||||||||||||||||| |||||| ||||||| ||||||||||||||||||| ||
Db 181 GTGCGGCCGTGCCGACCGCCGCAGCCGTGCCCATGTCAACCTCAACCTCAAGGACGAGCTGTTT 240

Qy 379 CCCAGCTGGGAGGGCCTTATTCTCGGGGCCCGAGGGGCCAGCGTGAAGCCCGGGGCCCGCATC 438
         |||||||||||||||||||||||||| ||||| | ||||||||||||||||||||||||||||
Db 241 CCCAGCTGGGAGGGCTCTGTTCTCAGGGTCGGCTGAGGGTCCGCTGAAGCCCGGGGCACGCATC 300

Qy 439 TTCTCTTTCGACGGCAGAGATGTCCTGACGCACCCCGGCCCCTGGCCCCGGAAGAGGCGTGTGG 498
         |||||||||||||||||| | ||||||||||||||||||| |||||||| |||||||||||||||
Db 301 TTCTCCTTGACGGCAAGGACGTCCTGAGGCACCCCACCTGGCCCCACCTGGCCCCCAGAAGAGCGTGTGG 360

Qy 499 CACGGCTCCGACCCCAGCGGGCGCGCCTGACCGAGCTGGCGGACG 558
         |||||||||| |||||||||||||||||||||||||||||||||||
Db 361 CATGGCTCGGACCCCAACGGGCGCAGGCTGACCGAGGTGTGAGACGTGGCGGACG 420

Qy 559 GAGGCCCCGGCGGCCACCGGGCGTCGTCCGCTGGCGGGCAGGCTGCTGGAGCAG 618
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 GAGGCTCCCTCGGCCACAGGGCCACCAGGCCTTCCTGCTCGTGCTGGGGGGCAGGCTCCCTGGGGCAG 480

Qy 619 GAGGCCGGCGAGCTGCGCCCACGCCTTCGTGGTGCTCTTGCATCGAGAACAGCGTCATGACC 678
         ||||||||||||||||||||| ||||| || ||||||||
Db 481 AGTGCCGCGAGCTGCCATCACGCCTACATCGTGCTCTGCATTGGAGAACAGCGTTCATGACT 540

Qy 679 TCCTTC 684
         ||||||
Db 541 GCCTCC 546

RESULT 10
US-10-292-418-3
; Sequence 3, Application US/10292418
; Publication No. US20030139365A1
; GENERAL INFORMATION:
; APPLICANT: Lo, Kin-Ming
; APPLICANT: Li, Yue
; APPLICANT: Gillies, Stephen D
; TITLE OF INVENTION: Expression and Export of Angiogenesis Inhibitors as
; TITLE OF INVENTION: Immunofusins
; FILE REFERENCE: LEX-006C1
; CURRENT APPLICATION NUMBER: US/10/292,418
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: 09/383,315
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: US 60/097,883
; PRIOR FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(549)
; OTHER INFORMATION: endostatin
US-10-292-418-3

Query Match   50.4%;   Score 418; DB 14;   Length 549;
Best Local Similarity 85.3%;   Pred. No. 1.8e-89;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 139 CACACCCACCAGGACTTCAGCTGGTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 198
         ||||||||||||||||||||||||||||| || | |||||||| |||||||||||
Db   1 CACAGCCACCGCGACTTCAGCGTGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA 60

Qy 199 GGGCGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGTGCTTCCAGCAGGCGCGCGCGG 258
         ||||||||||||||||||||||||||| ||||||||||| | ||||||||||||||||||
Db   61 GGCGGCATGCGGGGGCATCCGGGGGCGGACTTCCAGCAGTGCTTCCAGCAGGCGCGGGCCGTG 120

Qy 259 GGGCTGGCCGGCACCTTCCTGGGCCTTCGTGTCGCGGCTGCAGGACCTTCTACAGCATC 318
         ||||||||||||||||||||||||||||||||||||| | ||||||||| |||||||||
Db 121 GGGCTGGCGGGCACCTTCCCGCGCCTTCGTCCTGCCGCGGAGGACCTGTACAGCATC 180

Qy 319 GTGCGCCCGCGCCACCGGGGTGCCCGTCGTGTCAACCTCAGGGACGGAGGTGCTCTTC 378
         |||||||||||| ||||||||| |||||| |||||||||||||||||||||| || ||
Db 181 GTGCGGCCGTGCCGACCGCGGCAGCCGTGCCCATGCAACCTCAACCTCAAGGACGAGCTGTTT 240

Qy 379 CCCAGCTGGGAGGGCCTTATTCTCGGGGCTCCGAGGGCCTCAGCAGACGCGGGCCGATC 438
         ||||||||||||||||||||||||||| ||||| | ||||||||||||||||||||||||
Db 241 CCCAGCTGGGAGGGCTCTGTTCTCAGGGTCGGCTGAGGGTCCGCTGAAGCCCGGGGCACCGCATC 300

Qy 439 TTCTCTTTCGACGGCAGAGATGTCCTGACGCAGCACCCCGGCCCCTGGCCCCCGGAAGAGGCGTGTGG 498
         |||||||||||||||||| | ||||||||||| | |||||||||| |||||||||| |||||||||||||
Db 301 TTCTCCTTGACGGCAAGGACGTCCTGAGGCACCCCACCTGGCCCCACCTGGCCCCCAGAAGAGCGTGTGG 360

Qy 499 CACGGCTCCGACCCCAGCGGGCGCGCCTGACCGAGCAGCTACTGCGACGACGTGGCGGACG 558
         ||||||||| |||||||||||| ||||||||||||||| |||||||||||||||||||||||
Db 361 CATGGCTCGGACCCCAACGGGCGCAGGCTGACCGAGAGCTACTGTGAGACGTGGCGGACG 420

Qy 559 GAGGCCCCGGCGGCCACCGGGCGCAGGCGGTCGTCGGCTGCTGGAGCAG 618
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 GAGGCTCCCTCGGCCACAGGGCCACCAGGCCTTCCTGCTCGTGCTGGGGGGCAGGCTCCCTGGGGCAG 480

Qy 619 GAGGCCGGCGAGCTGCGCCCACGCCTTCGTGGTGCTCTGCATCGAGAACAGCGTCATGACC 678
         |||||||||||||||||||| ||||| || ||||||||
Db 481 AGTGCCGCGAGCTGCCATCACGCCTACATCGTGTCTCTGCATTGGAGAACAGCGTTCATGACT 540

Qy 679 TCCTTC 684
         ||||||
Db 541 GCCTCC 546

RESULT 11
US-09-873-676-30
; Sequence 30, Application US/09873676
; Patent No. US2002007289A1
; GENERAL INFORMATION:
; APPLICANT: MacDonald, Nicholas J.
; APPLICANT: Sim, Kim L.
; TITLE OF INVENTION: Angiostatin and Endostatin Binding Proteins and Methods of Use
; FILE REFERENCE: 05213-0378 (43170-259333)
; CURRENT APPLICATION NUMBER: US/09/873,676
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: US 60/209,065
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: US 60/289,387
; PRIOR FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-873-676-30

Query Match   50.4%; Score 418; DB 9; Length 552;
Best Local Similarity 85.3%; Pred. No. 1.8e-89;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy  139 CACACCCACCAGGACTTCCAGCTGCTGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 198
            ||||||||||| |||||||||||  || || ||||||||  | |||||||||| | ||
Db    1 CACAGCCACCGGCACTTCCAGCCGCCGTGCTTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA 60

Qy  199 GGCGGCATGCGAGGCATCCGGGAGGCAGCTTCCAGTGCTTCAGCACAGCGCGGCCGCGCG 258
            ||||||||||| ||||||||  ||| | || |||||||| |||||| ||||| |||||
Db   61 GGCGGCATGCGGGGCATCCGCGGGCCGCGGGCCGACTTCCAGCTGTTCCAGCAGGGCCGGCCGTG 120

Qy  259 GGGCTGCGCCGGCCACCTTCGGGGCCTTCCGTCGTCGCGGCTGCGAGGACCTTCACAGCATC 318
            |||||||  ||||||||||| | |||||  | | |  ||| |||  ||| |||||||||
Db  121 GGGCTGCGGGGCCACCTTCGCGCCTTCGCGCCGCCGTGTCCGCGGCCGCCTGTACAGCATC 180

Qy  319 GTGCGCGCCGCGACCGCACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 378
            |||||| | || |||| ||||| |||||| | ||  ||||| ||||||| ||| ||
Db  181 GTGCGCCGTGCCGACCGGACCGGCCGTGCCCATCGTGCACCTCAAGGACGAGCTGCTGTTT 240

Qy  379 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGCCAGCTGAAGCCCCGGCCGCCGCATC 438
            |||||||||||| ||  | |||| ||| | ||||| || || ||||||| |||| |||||
Db  241 CCCAGCTGGGAGGGCTCTGTTTCTCAGCTCTGAGGCTCCAAGGACCCCGGAGCCCGGCCACGCATC 300

Qy  439 TTCTCCTTCGACGGCCAGAGATGCTCCTGCAGCACCCCGGCCCGGCCGCCGAGAGGCGTGTGG 498
            ||||| || ||||||||  || || || || ||||||||   |||||   |||||| |||||
Db  301 TTCTCCTTTGACGGCCAAGGACGTCCTGAGGCACCCCACCTGGCCCCCAGAGAGAGCGGTGG 360

Qy  499 CACGGCTTCGACCCCGACGGGGCGGCGCCGCCGCGCCAGCTACTGCGAACGTCGGCGGACG 558
            || |||||||| |||| ||||||  || ||  | || || |||| |||||| ||||||||
Db  361 CATGGCTTCGGACCCCAAGGGGCGCCAGGGCTGCAGCGGCGTACTGAGAGGTGCGGCGGACG 420

Qy  559 GAGGCCCCGGCGGCCCACCGGGCAGGCGTCGTCGGTGCTCGTCGAGCAG 618
            |||||  ||| ||||||||||||||  | |   ||||||  |||||||
Db  421 GAGGCCTCCTCCGCCCAGGGGCCAGGGCGTCCTGGGGCAG 480

Qy  619 GAGGCGCGGCAGCTGCGCCACGCCTTCGTCGTGCTTCGTCGATCATCAGCGTCATGACC 678
            |||||| ||||||||| | |||||| |||||||| |  |||||||| |||| |||||
Db  481 AGTGCCCGCGAGCTGCCATCACGCCTACATCAGCGTTCAATTGAGAACAGCTTCATGACT 540

Qy  679 TCCTTC 684
            |||
Db  541 GCCTCC 546
```

RESULT 12
US-10-080-797-2
; Sequence 2, Application US/10080797
; Publication No. US20020183253A1

; GENERAL INFORMATION:
; APPLICANT: Campochiaro, Peter A.
; APPLICANT: Dixon, Katharine H.
; APPLICANT: Brazzell, Romulus K.
; TITLE OF INVENTION: METHOD FOR TREATING OCULAR
; TITLE OF INVENTION: NEOVASCULARIZATION
; FILE REFERENCE: 4-31881A
; CURRENT APPLICATION NUMBER: US/10/080,797
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 551
; TYPE: DNA
; ORGANISM: Human
US-10-080-797-2

Query Match   50.3%; Score 417; DB 13; Length 551;
Best Local Similarity 85.3%; Pred. No. 3.1e-89;
Matches 465; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy  140 ACACCCACCAGGACTTCCAGCTGCTGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 199
            ||||||||| |||||||||||  || || ||||||||  | |||||||||| | ||
Db    1 ACAGCCACCGGCACTTCCAGCCGCCGTGCTTCCACCTGGTTGCGCTCAACAGCCCCCTGTCAG 60

Qy  200 GCGGCATGCGAGGCATCCGGGAGGCAGCTTCCAGTGCTTCAGCACAGCGCGGCCGCGCG 259
            ||||||||| ||||||||  ||| | || |||||||| |||||| ||||| |||||
Db   61 GCGGCATGCGGGGCATCCGCGGGCCGCGGGCCGACTTCCAGCTGTTCCAGCAGGGCCGGCCGTG 120

Qy  260 GGCTGCGCCGGCCACCTTCGGGGCCTTCCGTCGTCGCGGCTGCGAGGACCTTCACAGCATG 319
            |||||  ||||||||||| | |||||  | | |  ||| |||  ||| |||||||||
Db  121 GGCTGCGGGGCCACCTTCGCGCCTTCGCGCCGCCGTGTCCGCGGCCGCCTGTACAGCATG 180

Qy  320 TGCGCGCCGCGACCGGCACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTCC 379
            |||| | || |||| ||||| |||||| | ||  ||||| ||||||| ||| ||
Db  181 TGCGCCGTGCCGACCGGACCGGCCGTGCCCATCGTGCACCTCAAGGACGAGCTGCTGTTTC 240

Qy  380 CCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGCCAGCTGAAGCCCCGGCCGCCGCATCT 439
            |||||||||| ||  | |||| ||| | ||||| || || ||||||| |||| |||||
Db  241 CCAGCTGGGAGGGCTCTGTTTCTCAGCTCTGAGGCTCCAAGGACCCCGGAGCCCGGCCACGCATCT 300

Qy  440 TCTCTTCGACGGCCAGAGATGTCCTGCAGCACCCCGGCCCGGCCGCCGAGAGGCGTGTGGC 499
            ||| || ||||||||  || ||| || || ||||||   |||||   |||||| ||||||
Db  301 TCTCCTTTGACGGCCAAGGACGTCCTGAGGCACCCCACCTGGCCCCCAGAGAGAGCGGTGGC 360

Qy  500 ACGGCTTCGACCCCGACGGGGCGGCGCCGCCGCGCCAGCTACTGCGGACGTCGGCGGACGG 559
            | |||||||| |||| ||||||  || ||  | || || |||| ||||| ||||||||
Db  361 ATGGCTTCGGACCCCAAGGGGCGCCAGGGCTGCAGCGGCGTACTGTGAGAGGTGCGGCGGACGG 420

Qy  560 AGGCCCCGGCGGCCCACCGGGCAGGCGGTCGTCGGTGCTCGGTGGAGCAGG 619
            |||||  ||| ||||||||||||||  |  |   ||||||   |||||||
Db  421 AGGCTCCTCCGCCCAGGGGCCAGGGCGTCCTGGGGCAGGTCCTGGGGGCAGGA 480

Qy  620 AGGCCGGAGCTGCGCCCACGCCTTCGTCGTGCTTCGTCATCGAGGTCATGACCT 679
            |||| | |||||| | |||||| |||||||| |  ||||| |||| |||||
Db  481 GTGCCCGCGAGCTGCCATCACGCCTACATCAGCGTTCAATTGAGAACAGCTTCATGACTG 540

Qy  680 CCTTC 684
            |||
Db  541 CCTCC 545
```

RESULT 13
US-10-131-241-51
; Sequence 51, Application US/10131241
; Publication No. US2003001279A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolife
; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241

```
;   CURRENT FILING DATE:   2002-07-22
;   PRIOR APPLICATION NUMBER: US 09/413,049
;   PRIOR FILING DATE:   1999-10-06
;   PRIOR APPLICATION NUMBER: US 09/316,802
;   PRIOR FILING DATE:   1999-05-21
;   PRIOR APPLICATION NUMBER: US 60/086,586
;   PRIOR FILING DATE:   1998-05-22
;   NUMBER OF SEQ ID NOS: 65
;   SOFTWARE: Patentin version 3.1
;   SEQ ID NO 51
;   LENGTH: 632
;   TYPE: DNA
;   ORGANISM: Murinae sp.
US-10-131-241-51

Query Match     50.1%;  Score 415.6;  DB 14;  Length 632;
Best Local Similarity   82.8%;  Pred. No. 6.6e-89;
Matches  523;  Conservative   0;  Mismatches  29;  Indels  80;  Gaps  1;

Qy   139 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 198
             ||||||||||||||||||||||||||||| |||||||||||||||||||| ||||||||-|
Db     1 CATACCCATCAGGACTTTCAGCCAGTGCTCCACCTGGTGCACTGAACACCCCCCTGTCT 60

Qy   199 GGCGGCATGCGAGGCATCCGGGAGGAGGGACTTCCAGTGCTTCAGCAGGCGGCGCGCCG 258
             |||||||||||||||||  |||| |||||||||||||||||||||||||||||||||||
Db    61 GGAGGCATGCGAGGCGTGGTATCCGTGGAGCAGATTTCCAGTGCTTCCAGCAGCCCGGTG 120

Qy   259 GGGCTGGCCGGCACGTTCCGGACCCTTCCGGGCCTTCCGTGTCGTCGGGC---------- 298
             |||||||||||||||||||||  || |||||| | || ||||||||| |||||
Db   121 GGGCGTGCGGGCACCTTCCGGGCTTTCCGGCTTTCCTGTCCTCTAGGCTGCGAGGATCTCTATAGAGCATC 180

Qy   299 -------------------------------------------------------------- 298
                                                                         
Db   181 GTGCGGCCGTGCTGACCGGGGGGTCTGTGCCCATCGTCAACCTGAAGGACGAGAGGTGCTATCT 240

Qy   299 TGCAGGACCCTTCTACAGCATGTGCCGCCGCACCGCACCGGGGTGCCCGTCGTCAACC 358
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TGCAGGACCCTCTACAGCATCGTGCCGCGCGCGCGCCGCACCGGGGTGCCCGTCGTCAACC 300

Qy   359 TCAGGGACGAGGTGCTTCCTCCCAGCTGGAGGCCTTATTCTCGGGCTTCGAGGGCCAGC 418
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TCAGGGACGAGGTGCTTCCTCCCAGCTGGAGGCCTTATTCTCGGGCTTCCGAGGGCCAGC 360

Qy   419 TGAAGCCCGGGGCCCGCATCTTCTTTGACGGCAGAGATGTCCTGCAGCACCCCGCCT 478
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGAAGCCCGGGGCCCGCATCTTCTCTTTGACGGCAGAGATGTCCTGCAGCACCCCGCCT 420

Qy   479 GGCCCCGGAAGAGCGTGTGGCACGCGTCCCAGCGGGCGCGCTCCAGCGACGACAGCT 538
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GGCCCCGGAAGAGCGTGTGGCACGCGTCCCAGCGGGCGCGCTCCAGCGACGACAGCT 480

Qy   539 ACTGCGAGACGTGGCGACGCAGGAGGCCCCGGCGAGCTGCGCGTCGTCGCTGCTGG 598
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ACTGCGAGACGTGGCGACGCAGGAGGCCCACGGAGCAGGCGTCGTCGCTGCTGG 540

Qy   599 CGGGCAGGCTGCTGGAGACGAGGAGGCGGAGCTGCCCGAGCTGCCACGCCTTCGTGGTGTCTGCA 658
             ||||||||||||||||||||||||||||||||||||||||||||
Db   541 CGGGCAGGCTGCTGGAGACGAGGAGGCGGAGCTGCCCGAGCTGCCACGCCCTTCGTGCTGTCTGCA 600

Qy   659 TCGAGAACAGCGTCATGACCTTCCTTCCCAAG 690
             ||||||||||||||||||||||||||||||||
Db   601 TCGAGAACAGCGTCATGACCTTCCTTCCCAAG 632
```

RESULT 14
US-10-210-172-161

```
Query Match     49.7%;  Score 411.6;  DB 12;  Length 555;
Best Local Similarity   84.6%;  Pred. No. 5.8e-88;
Matches  462;  Conservative   0;  Mismatches  84;  Indels   0;  Gaps   0;

Qy   139 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCCCTGAACAGCCCGCAGCCG 198
             ||||  |||  ||||||||||||||||||  |||||||||  ||  |||||  |||||| ||
Db     4 CACAGCCACCGGGACTTCCAGCGACTTCCACCTGGTTGCGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA 63

Qy   199 GGCGGCATGCGAGGCGACTTCCGGGGAGGGGAGCGGACTTCCAGTGCTTCCAGTGCTTCCAGCGCGCGCG 258
             ||||||  |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    64 GGCGGCATGCGGGGCATCCGGGCGGGGGGCATCCGGGGCGGGGGCCGACTTCCAGTGCTTCCAGCAGCCGGCGGGGCGGGCGGGCGTG 123
```

```
Db       ||||||||| |||||| ||||||||| || ||||||| |||| | ||||||||||||||||||
      1 GACTTCCAGCCGGTGCTGCCACCTGGTTGCGCTCAACAGCCCCTGTCAGGCGGCATGCGG   60

Qy  211 GGCATCCGGGACGGGACTTCCAGTGCTTCCAGGTGGCGGCTCCACGAGGCGGCGGCCGGC  270
        ||||||||||||||| ||||||||||||||| |||| ||||||||||||||||| |||
Db   61 GGCATCCGCGGGGCCGACTTCCAGTGCTTCAGCAGCGGCGGCCCGTGGGGCTGGCCGGGGC  120

Qy  271 ACCTTCCGGCCCTTCCTGTGCTCGCGGCTGCAGGACCTCACAGCATCGTGCGCGCGCCGCC  330
        ||||||||||||||||||||||||||||||| | |||||||||||||||||||||||||
Db  121 ACCTTCCGCGCCCTTCCTGTCCTCGCGGCTGCAGGACCTCGTACAGCATCGTGCGCCGTGCC  180

Qy  331 GACCGCACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTCCCCAGCTCGGGAG  390
        |||||||||||||| |||||||||| |||| ||||||||||||||||||||||||||||
Db  181 GACCGCGCAGCCGTGCCCATGCTCAACCTCAGGACGACGAGCTGCTCGTTTCCCAGCTCGGGAG  240

Qy  391 GCCTTATTCTCGGGCTCCGAGGCGCCAGCTGAAGCCCGGGCCCGCATCTTCTTCTTTCGAC  450
        |||  |||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db  241 GCTCCGTGTTCTCAGGCTTCAGGGTCCCTGAGGGTCCGCTGAAGCCCGGGCACGCCATCTTCTCCTTTGAC  300

Qy  451 GGCAGAGATGTCCTGCAGCACCCCGCTGGGCCCGGAAGAGCCGTGTGGCACGGCTCCGAC  510
        |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GGCAAGGAGGACGTCCTGAGGCACCCCACCTGGCCCCCAGAGAGAGCGTGTGGCATCGGCTCCGAC  360

Qy  511 CCCAGCGCGGCGGCCGCCGCCTACCGCAGCTACTGCGAGACGTGCGGCGACGGAGCCCCGCG  570
        |||||||||||||||||||||||||  ||||||| |||||||||||||||||||||||||
Db  361 CCCAACGGGGGCGCAGGCTGACCGAGAGCTACTGTGACCGAGGCTACTGCGAGACTGGGCGCAGC  420

Qy  571 GCCACCGGGCAGGCGTCGTTCGCGGTGCCAGGCTGCTCGGAGCAGGAACACGGCCGCGAGC  630
        |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db  421 GCCACGCGGCCGCCAGGCCTCCTCGCGCGCTGCTCGTGGGGGGGGCAGGCAGGGCTCCGGCGGAGTGCCGCGAGC  480

Qy  631 TGCCGCCACGCCTTCGTCGGTGTGTGCTCGCATCGAGAACAGCGTCATGACCTCCTTC  684
        ||||||||| ||||||||| | |||||||||| ||||| | ||||||||||||
Db  481 TGCCATCACGCCTACATCGTGCTTCCTGCATTGAGAACAGCTTCATGACTGCCTCC  534

Search completed: March 29, 2004, 16:40:25
Job time : 340.85 secs
```

```
Qy  259 GGGCTGGCCGGCGGCACCTTCCGGGGCCTTCCTGTGCTCGCGGCTGCAGGACCTCTACAGCATC  318
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  124 GGGCTGGGCGGGGCACCTTCCGCGCCCTTCCTGTCCTCGCGGCTGCCTTCGCGGCTGCAGGACCTCGTACAGCATC  183

Qy  319 GTCGCGCGCGGCCGACCCACGGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTCC  378
        ||||||  ||||| |  ||   |||||||||||  ||  ||||||||||||||||||||||
Db  184 GTGCGCCGTGCCGACCGCGCAGCCGTGCCCATGCTCAACCTCAAGGACACGAGCTGCTGCTGTTT  243

Qy  379 CCCAGCTCGGGAGGCCTTATTCTCGGGCTCCGAGGCGGGCCAGCTGAAGCCCGGGGCCCGCATC  438
        |||||||||||||||| |||||||||||||||||  |||||||||||||||||||||||||
Db  244 CCCAGCTCGGGAGGCTCTGTTCTCCAGGCTCCTGAGGGTCCGCTGAAGCCCGGGGGCACGCATC  303

Qy  439 TTCTTTCTGACGGCAGAGATGTCCTGCAGCACGGGCCCGCTGGGCCCGGAAGAGCCGTGTGG  498
        ||||  ||||||||| ||||||||||||||||||||||||  ||||||||||||||||||
Db  304 TTCTCCTTTAACGGCAAGGAGGACGTCCTGAGGCACCCCACCTGGCCCCCAGAGAGCCGTGTGG  363

Qy  499 CACGGCTCCGACCCCAGCGCGGCCGCCGCCTACCGCAGCTACTGCGAGACGTGCGGCGGACG  558
        |||||||||||  ||||||||||||  |||||||||||||||||  | ||||||||||||||
Db  364 CATGGCTCGACCCCAACGGGGCCAGCGGCGTGCCACCCCACCTGGCCTGCTGCTGCTGCGGACG  423

Qy  559 GAGGCCCCGGGGCACCGGGCACGCGTCGTTCGCGGTGCTGCTGGAGCAG  618
        ||||||||  |||||||||||||||||||||||||||||||||||||||
Db  424 GAGGCTCCCTCGGCCACGGGCCACGGCCCAGGGCCACGCTCCTGCTGCTCGCTCCTGGGGCAG  483

Qy  619 GAGGCCGCCGAGCTGCCGCCACGCCTTCGCGGTGCTCTGTGGTGCTCTGCATCGAGAACAGCGTCATGACC  678
        |||| |||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db  484 AGTGCCGCGAGCTGCCATCACGCCTACATCGTGCTTCCTGCTTGAGAACAGCTTCATGACT  543

Qy  679 TCCTTC  684
        ||||||
Db  544 GCCTCC  549
```

RESULT 15
US-10-042-347-6
; Sequence 6, Application US/10042347
; Publication No. US20030114370A1
; GENERAL INFORMATION:
;; APPLICANT: O'Reilly, Michael S.
;; APPLICANT: Folkman, M. Judah.
;; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Endostatin Protein and Peptide Fr
;; TITLE OF INVENTION: Thereof
;; FILE REFERENCE: 05213-0880   (43170-249874)
;; CURRENT APPLICATION NUMBER: US/10/042,347
;; CURRENT FILING DATE: 2002-01-11
;; PRIOR APPLICATION NUMBER: US 09/315,689
;; PRIOR FILING DATE: 1999-05-20
;; PRIOR APPLICATION NUMBER: US 60/106,343
;; PRIOR FILING DATE: 1998-10-30
;; PRIOR APPLICATION NUMBER: US 09/154,302
;; PRIOR FILING DATE: 1998-09-16
;; PRIOR APPLICATION NUMBER: US 08/740,168
;; PRIOR FILING DATE: 1996-10-22
;; PRIOR APPLICATION NUMBER: US 60/005,835
;; PRIOR FILING DATE: 1995-10-23
;; PRIOR APPLICATION NUMBER: US 60/023,070
;; PRIOR FILING DATE: 1996-08-02
;; PRIOR APPLICATION NUMBER: US 60/026,263
;; PRIOR FILING DATE: 1996-09-17
;; NUMBER OF SEQ ID NOS: 6
;; SOFTWARE: Patentin version 3.1
; SEQ ID NO 6
;; LENGTH: 534
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-042-347-6

Query Match   49.6%; Score 410.8; DB 14; Length 534;
Best Local Similarity   85.6%; Pred. No. 9e-88;
Matches 457; Conservative   0; Mismatches 77; Indels   0; Gaps   0;

Qy 151 GACTTCCAGCTGGTGCTGCACCTGGTGCCCTGGTGGCCCTGAACAGCCCCGACGCCCGGCGCCATGCGA 210

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 11:07:01 ; Search time 2546.3 Seconds
                                      (without alignments)
                                      9722.244 Million cell updates/sec

Title: US-09-938-391-1
Perfect score: 829
Sequence: 1 ccctgcgggcagatgacat.........aataaaggaagccaaagag 829

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 1493109076 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :   EST:*
          1: em_estba:*
          2: em_esthum:*
          3: em_estin:*
          4: em_estmu:*
          5: em_estov:*
          6: em_estpl:*
          7: em_estro:*
          8: em_htc:*
          9: gb_est1:*
         10: gb_est2:*
         11: gb_htc:*
         12: gb_est3:*
         13: gb_est4:*
         14: gb_est5:*
         15: em_estfun:*
         16: em_estcom:*
         17: em_gss_hum:*
         18: em_gss_inv:*
         19: em_gss_pln:*
         20: em_gss_vrt:*
         21: em_gss_fun:*
         22: em_gss_mam:*
         23: em_gss_mus:*
         24: em_gss_pro:*
         25: em_gss_rod:*
         26: em_gss_phg:*
         27: em_gss_vrl:*
         28: gb_gss1:*
         29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Match Length	DB ID	DB Length	Description
1	503.2	60.7	4230	11	BC063833	BC063833 Homo sapi
2	493.2	59.5	881	14	CD105862	CD105862 AGENCOURT
3	459.6	55.4	929	13	BQ672290	BQ672290 AGENCOURT
4	451	54.4	979	13	BQ673186	BQ673186 AGENCOURT

5	445.2	53.7	551	10	BF074459	BF074459 221883 MA
6	441.6	53.3	832	12	BG387051	BG387051 60245474 9
7	436.8	52.7	4192	11	BC062931	BC062931 Mus muscu
8	428	51.6	843	10	BF385854	BF385854 60204602 1
9	425.8	51.4	715	9	AU125614	AU125614 AU125614
10	425.8	51.4	947	13	BU556872	BU556872 AGENCOURT
11	422.8	51.0	707	10	BE908201	BE908201 60150045 8
12	422.4	51.0	757	10	BE906253	BE906253 60150223 7
13	417	50.3	874	12	BI412588	BI412588 60299046 8
14	413	49.8	504	14	CF789984	CF789984 867322 MA
15	401.6	48.4	657	14	CB444165	CB444165 695295 MA
16	398.2	48.0	944	13	BU859398	BU859398 AGENCOURT
17	394.4	47.6	703	13	BU615520	BU615520 UI-H-FG0-
18	387	46.7	683	12	BM683067	BM683067 UI-E-EO1-
19	386.6	46.6	682	9	AW089583	AW089583 xd20f03.x
20	385.6	46.5	735	14	CF728236	CF728236 UI-M-HB0-
21	379.6	45.8	618	9	AV696242	AV696242 AV696242
22	379.2	45.7	715	14	CB596713	CB596713 AGENCOURT
23	378.2	45.6	897	10	BI080524	BI080524 60287700 5
24	376.8	45.5	611	10	AW911243	AW911243 ur83h10.y
25	376.6	45.4	782	9	AI326391	AI326391 mm18f09.x
26	374.4	45.2	884	12	BI161007	BI161007 60286521 3
27	373.8	45.1	753	12	BI904605	BI904605 60316841 1
28	373.4	45.0	835	12	BI526580	BI526580 60293545 4
29	368.6	44.5	723	12	BI247582	BI247582 60296004 1
30	366.6	44.2	1093	13	BQ723254	BQ723254 AGENCOURT
31	364.6	44.0	668	14	CF724654	CF724654 UI-M-GZ0-
32	356.8	43.0	915	10	BF166139	BF166139 60177658 6
33	356.6	43.0	664	11	BU632049	BU632049 UI-H-FE1-
34	354.2	42.7	650	9	AI858615	AI858615 w140f01.x
35	353.8	42.7	846	13	BU540812	BU540812 AGENCOURT
36	353	42.6	720	12	BI147444	BI147444 60291400 8
37	352.6	42.5	639	12	BM998137	BM998137 UI-H-DT1-
38	350	42.2	634	13	BQ772348	BQ772348 UI-H-E21-
39	350	42.2	660	10	AW192502	AW192502 x145e10.x
40	349.4	42.1	652	13	BU352506	BU352506 60352798 2
41	348.8	42.1	747	12	BG967333	BG967333 60283364 9
42	348.4	42.1	587	9	AA288198	AA288198 vb15b10.r
43	348.2	42.0	634	9	AI970297	AI970297 wr09c02.x
44	343.8	41.5	745	10	AW243446	AW243446 xm95d11.x
45	343.6	41.4	690	12	BI219399	BI219399 60293675 6

ALIGNMENTS

RESULT 1
BC063833
LOCUS   BC063833   4230 bp   mRNA   linear   HTC 09-DEC-2003
DEFINITION Homo sapiens cDNA clone IMAGE:6181818, containing frame-shift
            errors.
ACCESSION   BC063833
VERSION   BC063833.1 GI:39645297
KEYWORDS   HTC.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 4230)
AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,

```
          Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
          Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
          Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
  TITLE   Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
          Generation and initial analysis of more than 15,000 full-length
          human and mouse cDNA sequences
  JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  PUBMED  12477932
REFERENCE 2  (bases 1 to 4230)
  AUTHORS Strausberg,R.
  TITLE   Direct Submission
  JOURNAL Submitted (08-DEC-2003) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
          USA
  REMARK  NIH-MGC Project URL: http://mgc.nci.nih.gov
  COMMENT Contact: MGC help desk
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Dr. James R. Lupski
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Genome Sequence Centre,
          BC Cancer Agency, Vancouver, BC, Canada
          info@bcgsc.bc.ca
          Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
          Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
          Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
          Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
          Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
          Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
          Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
          George Yang, Scott Zuyderduyn, Marco Marra.

          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
          Series: IRAK Plate: 141 Row: c Column: 6
          This clone was selected for full length sequencing because it
          passed the following selection criteria: matched mRNA gi: 18765747
          This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1..4230
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6181818"
                     /tissue_type="Peripheral Nervous System, dorsal root
                     ganglion"
                     /clone_lib="Lupski_dorsal_root_ganglion"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
ORIGIN

Query Match      60.7%;  Score 503.2; DB 11;  Length 4230;
Best Local Similarity  77.6%;  Pred. No. 3.8e-63;
Matches  664; Conservative  0; Mismatches  153;  Indels  39; Gaps  3;

Qy      1 CCCTGCGGGCAGAAGAGATGACATCCTGGCCCGCCGCCCCCGCCCTGCTGACCCCAGCCTAC 60
            ||||| |||||| | |||| |||||||||| |||||||| ||||||||||| ||||||||||
Db   3369 CCCTGGCGGGCAGAGTGACATCCTGGCCAGCCCGAGCCCTGCCCGCTGACCCCAGCCTAC 3428

Qy     61 CCCGGGGCCCCGACCACGGCTCCTACGGTGCACTTCCAGCGGCTCGCCCACTGGTGGG 120
            |||||||| ||| ||||||| |||||||| |||||||||||||||||||||||||| 
Db   3429 CCCGGAGCCCCCGCCACCAGCTCCTACGGTCACCTGCGGCCGCAGCACACCCACAG---- 3484

Qy    121 CCCGTCCACACCCACACCCACCACCCACAGGACTTCCAGCGGCTGGTGCTGCACCTGGTGGCC 180
            |||||| ||||||||||||||||||||||||||||| ||||||||||||||||||||||| 
Db   3485 ------CCCACCCGCCACCAGCCACCCGCGACTTCCAGCCGTGCTCCACCTGGTTGCG 3536

Qy    181 CTGAACAGCCCCAGCCGCGGGGGCCGCATGCGCAGCCGGGGAGGCCGGACTTCCAGTGCTTC 240
            ||||||||| |||||||| |||||||||||||||||||| |||||||||||||||| 
Db   3537 CTCAACAGCCCCCCTGTCAGCGGGCATGCGGGGGCATCCGGGCGGGCCCACTTCCAGTGCTTC 3596

Qy    241 CAGCAGGCGCGCGCGCGCGCGCGGGGCTGGCCCGGCACCTTCCGTGCTGTGTCGCGGCTG 300
```

```
Db   3597 CAGCAGGCGCGGGGCCCTGGGGGTGCGGCGGCTGGGGGCACTTCCGCGGCCGCCTTCCGCGTGCTGTCCTCGTCGCCGCCTG 3656
Qy    301 CAGGACCTTCTACAGCATCGTGCGCGGCCGCGCGGGCCGCACCGGGGGTGCCCGTGCTGTCAACCTC 360
Db   3657 CAGGACCTGTACAGACATCGTCGCGGCCGCGGCACCCGGGGCCGCACCGGGGGTGCCCGTGCCCATCGTCAACCTC 3716
Qy    361 AGGGAGGAGGTGCTCTTCCCCAGTGGGAGGCCTTATTCTCGGGGCTCCGAGGGCCAGCTG 420
Db   3717 AAGGACGAGCTGCTGTTTCCCAGCTGGGGAGGGCTCTGTTCTCTCTGTTCTCTCGGGGTCCGTG 3776
Qy    421 AAGCCCGGGCCGCCGCCATCTTCTCTTCGACGGCAGAGATGTCCTGCAGCACCCGCCTGG 480
Db   3777 AAGCCCGGGGCACCGGCACCATCTTCTCCTTGACGGCAAGGACGTCCTGAGGCACCCACCTGG 3836
Qy    481 CCCCGGAAGAGAGCGTGTGGCACGGCTCCCCACCCCAGCGGCGGCGGCGCCTGACCCAGCAGCTAC 540
Db   3837 CCCCAGAGAGACGGTGTGGCATGGCGTCGGACCTCGGGACCCCACGGGGCGCAGGCTCACCGGAGCTAC 3896
Qy    541 TGCGAGACGTGGCGGACGAGGGCCGCGGCCAGCGTGCCCGTCGCTGCTGGCG 600
Db   3897 TGTGAGACGTGGCGGACGAGGGCTTCCCTGGCCACGGCGGCCAGCCTCCTCGCTGCTGGGG 3956
Qy    601 GGCAGGCTGCTGGAGCAGGAGGGCCGCGAGCTGCCGCCACGCCTTCGTCGTGTGTCGCATC 660
Db   3957 GGCAGGCTCCTGGGCGCAGAGTGCGGCGGAGCTGCCCATCACCGCCTACACCGTGCTCTGCATT 4016
Qy    661 GAGAACAGCGTCATGACCTCCTTCTCCAAGTAGGGCCCGCGGCCCACCGGACAGGCGGGG 720
Db   4017 GAGAAACAGCTTCATGACTCGCTCCAAGTAGCCACCGCCTGGATGCAGATGGCCGGAGAGG 4076
Qy    721 GAGGGGGCGCCCGCAGGAGAGATCACGCGCCGCGCCGGGG-------------------- 756
Db   4077 ACCGGCGGGCTCGGAGGAGGAAGCCCCACCGTGGGCAGGGAGCAGCGGCCGCGGCCCCTGGCCC 4136
Qy    757 --GGGCTGGCCGGGACGCGTTGGCTCGCACCG-TCACGTTTAATGTAATCCTCAAGAAATA 813
Db   4137 CAGGACCTGGTGCCATACTTTCCTGTATAGTTCAGTTCATGTAATCCTCAAGAAATA 4196
Qy    814 AAAGGAAGCCAAAGAG 829
Db   4197 AAAGGAAGCCAAAGAG 4212
```

RESULT 2
CD105862
CD105862
LOCUS   CD105862   881 bp   mRNA   linear   EST 15-MAY-2003
DEFINITION AGENCOURT_14021788 NIH_MGC_179 Homo sapiens cDNA clone
            IMAGE:30365831 5', mRNA sequence.
ACCESSION   CD105862
VERSION   CD105862.1   GI:30759036
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 881)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
            cDNA Library Preparation: Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM424 row: g   column: 24
            High quality sequence stop: 689.
FEATURES   Location/Qualifiers
     source   1..881

Right column

```
ACCESSION   BQ672290
VERSION     BQ672290.1  GI:21783124
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 929)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
  TITLE     Unpublished (1999)
  JOURNAL
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2458  row: c   column: 06
            High quality sequence stop: 588.
FEATURES             Location/Qualifiers
     source          1..929
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6275261"
                     /tissue_type="epidermoid carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_102"
                     /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of California, Berkeley) using ZAP-cDNA
                     synthesis kit (Stratagene) and Superscript II RT (Life
                     Technologies). Note: this is a NIH_MGC Library."

ORIGIN
Query Match              55.4%;  Score 459.6;  DB 13;  Length 929;
Best Local Similarity    77.4%;  Pred. No. 6.4e-57;
Matches  602;  Conservative   0;  Mismatches 159;  Indels  17;  Gaps   3;

Qy     1 CCCTGGCGGGCAGATGACATCTCGCCGCGCCCCCGGCCTGTGGACCCCAGCCCTAC 60
           |||||||||||||||||||||||||||||||||    |||| ||||||||||||||
Db   101 CCCTGGCGGGCAGATGACATCTCGCCGCGCCCCCTGCCCGGCTCGCCCCACTGGTGG 160

Qy    61 CCCGGGGCCCCGACCACCAGGCTCCTACGTGCACTTCAGCCGGCGGCCGCCCACTGGTGGG 120
           ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   161 CCCGGAGCCCCGACCACGCTCCTACGTGCACTTCAGCCGGCGCCGCCCGACCCACCAAG---- 216

Qy   121 CCGTCCACACCCACCACCCACCAGGACTTCCAGTCGTGCTGCACCTGGTGGCC 180
           |||||||||||||||||| ||||||||||||||||| |||||||||||||||
Db   217 -------CCCACCCGCCCACAGCCACCCGACCCGTCCAGCCGTCCACCTGGTTGCG 268

Qy   181 CTGAACAGCCCGCAGCCGGCGGGGCATGCGAGGCATCCGGGAGCGGACTTCCAGTGCTTC 240
           ||||||| |||||| |||||||||||||||||||||||||||||||||||||||||||
Db   269 CTCAACAGCCCCGTGTCAGCGGGGCATGCGGGGCATCCGCGGGGGCGACTTCCAGTGCTTC 328

Qy   241 CAGCAGGGCGGCGGCCCGCGGCGGGCTGCGCCGGCACTTCCTGGCCTTCCTGTCGTCGCGGCTG 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   329 CAGCAGGGCGGCGGCCCGCGGCGGGCTGCGCCGGCACTTCCTGGCGGCCTTCCTGTCGTCGCGGCTG 388

Qy   301 CAGGACCTCTACAGCATCGTGCGCGCGGCACCGGGGTGCCCGTGCTCAACCTC 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   389 CAGGACCTCTACAGCATCGTGCGCGCGGCACCGGGGTGCCCATCGTCAACCTC 448

Qy   361 AGGGACGAGGTGCTTCTCCCCAGCTGGGAGGGCCTTATTCTCGGGCTCCAGCTG 420
           |||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   449 AGGACGAGGTGCTGTTTCCCAGCTGGGAGGGCCTTGTTCTCAGGCTCTGAGGGTCCGCTG 508
```

Left column

```
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:10365831"
                     /tissue_type="Pituitary"
                     /lab_host="DH10B-TonA ( T1 and T5 phage resistances)"
                     /clone_lib="NIH_MGC_179"
                     /note="Organ: brain; Vector: pCMV-SPORT6.1; Site_1: EcoRV
                     (destroyed); Site_2: NotI; Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed upon
                     cloning). Average insert size 1.1 kb. Library was
                     constructed by (Invitrogen). Note: this is a NIH_MGC
                     Library."

ORIGIN
Query Match              59.5%;  Score 493.2;  DB 14;  Length 881;
Best Local Similarity    83.9%;  Pred. No. 8.9e-62;
Matches  574;  Conservative   0;  Mismatches  98;  Indels  12;  Gaps   1;

Qy     1 CCCTGGCGGGCAGATGACATCTCGCCGCGCCCCCGGCCTGTGGACCCCAGCCCTAC 60
           |||||||||||||||||||||||||||||||||    |||| |||||||||||||||
Db   166 CCCTGGCGGGCAGATGACATCTCGCCGCGCCCCCTGCCCGGCTCGCCCCACTGGTGGG 225

Qy    61 CCCGGGGCCCCGACCACCAGGCTCCTACGTGCACTTCAGCCGGCGGCCGCCCACTGGTGGG 120
           ||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   226 CCCGGAGCCCCGACCACGCTCCTACGTGCACTTCAGCCGGCGCCGCCCGACCCACCAAG---- 281

Qy   121 CCGTCCACACCCACCACCCACCAGGACTTCCAGTCGTGCTGCACCTGGTGGCC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   282 -------CCCACCCGCCCACAGCCACCCGACCCGTCCAGCCGTCCACCTGGTTGCG 333

Qy   181 CTGAACAGCCCGCAGCCGGCGGGGCATGCGAGGCATCCGGGAGCGGACTTCCAGTGCTTC 240
           ||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   334 CTCAACAGCCCCGTGTCAGCGGGGCATGCGGGGCATCCGCGGGGGCGACTTCCAGTGCTTC 393

Qy   241 CAGCAGGGCGGCGGCCCGCGGCGGGCTGCGCCGGCACTTCCTGGCCTTCCTGTCGTCGCGGCTG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   394 CAGCAGGGCGGCGGCCCGCGGCGGGCTGCGCCGGCACTTCCTGGCGGCCTTCCTGTCGTCGCGCGTG 453

Qy   301 CAGGACCTCTACAGCATCGTGCGCGCGGCACCGGGGTGCCCGTGCTCAACCTC 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   454 CAGGACCTCTACAGCATCGTGCGCGCGGCACCGGGGTGCCCATCGTCAACCTC 513

Qy   361 AGGGACGAGGTGCTTCTCCCCAGCTGGGAGGGCCTTATTCTCGGGCTCCAGCTG 420
           |||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   514 AGGACGAGCTGCTGTTTCCCAGCTGGGAGGGCTTCTCAGGCTCTGAGGGTCCGCTG 573

Qy   421 AAGCCCGGGCCCCATCTCTTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCCTGG 480
           ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   574 AAGCCCGGGCCCCATCTCTTTCTCCTTTGACGGCAAGGACGTCCTGAGGCACCCCACCTGG 633

Qy   481 CCCGGAAGAGCGGTGTGGCAGGCGTGGCACGGGCTCCGACCCGCGGGGCGGACGACAGCTAC 540
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   634 CCCAGAAAGAGCGGTGTGGCATGGCGTGGCTGGGGACCCCAACGGGCGCAGGGCTGACCGAGAGCTAC 693

Qy   541 TGCGGAGACGTGGCGGACGGAGCGGAGGCCCGGCGGGCCAGGCGTGGTGCGCGGCGTGGCGG 600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   694 TGTGAGACGTGGCCGGACGGAGGGGCTCCCTCGGCCCAGGCCTTCGCGCGGCTGGGG 753

Qy   601 GGGAGGCTGCTGGAGCAGGAGGGCCGGAGCTGCGCACGCGGGCTTCGTCGGTGGTCCTCGCATC 660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   754 GGCAGGCTCCTGGGGGCAGGTGCCGCGAGCTGCCATCACGCCTACAATCGTGCTCTGCATT 813

Qy   661 GAGAACAGCGTCATGACCTCCTTC 684
           |||||||||||||| |||||||
Db   814 GAGAACAGCTTCATGACTGCCTCC 837

RESULT 3
BQ672290
LOCUS       BQ672290       929 bp   mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8354946 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6275261
            5', mRNA sequence.
```

Matches 606; Conservative 0; Mismatches 145; Indels 18; Gaps 5;

```
Qy      1 CCCTGGCGGGCAGATGACATCCTGGCCGGCCCCCGGCTGCTGGACCCCAGCCCTAC   60
            |||||||||||| ||| || ||||||||||||||||||| |||||||||||||||||
Db    101 CCCTGGCGGGCAGATGACATCCTGGCCGGCCCCCGGCTGCTGGACCCCAGCCCTAC  160

Qy     61 CCCGGGGCCCCGCACCACGGCTCCTACGTGCACTTCCAGCGGCTGCCCCACTGGTCGGG  120
            ||||| |||||| |||| |||||||||||||||||||||||||||||||||
Db    161 CCCGGAGCCCGCACCACAGCTCCTACGTGCACTTGCACCTGCGGCGGCGGGACCCACAAG----  216

Qy    121 CCGTCCACCACCACCACCACCCACCAGGACTTCCAGCTGCTGCACCTGGTGCC  180
            ||||||||||||||||||||||||||||||||||||||||
Db    217 --------CCCACCCGCCCCACAGCCACCCGGCACTTCCAGCCCGGTGCTCCACCTGGTTGCG  268

Qy    181 CTGAACAGCCCGCAGCCGGGCATGCGAGCGGCATCCGGGGAGCATCCGGGGGACTTCCAGTGCTTC  240
            |||||||||||||||||||||||||||||||||||||
Db    269 CTCAACAGCCCCGTGCAGCCGGGCATGCGGGGCATCCGGGGCATCCGGGGGACTTCCAGTGCTTC  328

Qy    241 CAGCAGGCGGCGGCCGGGGGCTGGCCGGCACCTTCCGGGCCTTCCTGTCGTCGCGGCTG  300
            ||||||||||||||||||||||||
Db    329 CAGCAGGCGGCGGGCCCGTGGGGCTGCCGGGCACCTTCCGGGCACCTTCCTGTCCTCGCGGCTG  388

Qy    301 CAGGACCTTACAGCATCGTGCGGCCGCCGCACCGGGGTGCCCGTGCCCATCGTCACCTC  360
            ||||||||||||||||||||||||||||||
Db    389 CAGGACCTGTACAGCATCGTGCGGCCGCCGCACCGGGGTGCCCGTGCCCATCGTCACCTC  448

Qy    361 AGGGACGAGTGCTCTTCCCAGCTGGGAGGCCTTATTCTCCGGCTCCGAGGGCCCAGCTG  420
            ||||||||||||||||||||
Db    449 AAGGACGAGTGCTGTTTCCCAGCTGGGAGGCCTCGTTTCTCAGGCTCAGGGTGAGGGTCCGCTG  508

Qy    421 AAGCCCGGGGCCCGCATCTTCTCTTTCGACGGGCAGAGATGTCCTGCAGAGATGTCTGAGACACCCGGCTGG  480
            ||||||||||||||||||||||||||||
Db    509 AAGCCCGGGGCCCGCACGCATCTTCCTTTAACGGCAAGGACGTCCTGAGGCACCCACCTGG  568

Qy    481 CCCCGGAAAGAGCGTGTGGCGACACGGCTCCAGCCCCCAGCCGGCGCCGCCGCCGCTCCGACCGACAGCTAC  540
            ||||||||||||||||||||||||||||
Db    569 CCCCAGAAAGAGCGTGTGGCATGGCGCGAGTGTCGTCGGCAGGCTGACCGGGAGAGCTAC  628

Qy    541 TGCGAGACGTTGGCGACGGAGGGCCCCGGCGGCCACCCGGGCAGGGCGTGTCGCTGGTGGGG  600
            |||||||||||||||||||
Db    629 TGTGAGACGTTGGGCGACGGAGGGCCCCGGCCACGGGCACCGAGGCCTCCTCCGGGGGGG  688

Qy    601 GGCAGG----CTGCTGGAGCAGGAGAGGGCGGCCGAGCTGCCGCCGCCCTTCGTGGTGCTCTGC  657
            ||||||||||||||||||||
Db    689 GGGCAAGCTCCCTGGGGCGGAGAGTGCCGGGCATGCCGCCATCACGCGCCTCACACTGTGGTCT  748

Qy    658 ATCGAGAAACAGCGGTCATGACCTCCTTCTCCCAAG--TAGGGCCGGGGCCCCACCGGACCAGG  715
            ||||||||||||||||
Db    749 GCATGACGCGGCTCATGACTGCCCGTC-CAAGAAACCACCCGCTGGAATGCGGAATGG  807

Qy    716 CGGGGGAGGGCGGGGGCCGCCCGAGGAGCCATCCGCCCGCCCCGGGGGGGCCTGG  764
            ||||||||||||||||||||
Db    808 CCCGGAAAGGACCGGCGGCGGCTCCGAAGGAGCCCCCACCCTGGGCAGGG  856
```

```
Qy    421 AAGCCCGGGGCCCGCATCTTCTCTTTCGACGGGCAGAGATGTCCTGCAGAGCACCCGGCTGG  480
            ||||||||||||||||||||
Db    509 AAGCCCGGGGCCCGGCACGCATCTTCTCCTTTAACGGGCAAGGACGTCCTGAGGCACCCACCTGG  568

Qy    481 CCCCGGAAAGAGCGTGTGGCGACGGAGGGCCCCGGCGGCCACCCGGGCAGGGCGTGTCGCTGGTGGGG  540
            ||||||||||||||||||||
Db    569 CCCCAGAAAGAGCGTGTGGCATGGCGCGAGTGTCGTCGGCAGGCTGACCGGGAGAGCTAC  628

Qy    541 TGCGAGACGTTGGCGACGGAGGGCCCCGGCCACGGGCAGGGCGTGTCGCTGGTGGGG  600
            ||||||||||||||||||||
Db    629 TGTGAGACGTTGGGCGACGGAGGGCCCCGGCCACGGGCACCGAGGCCTCCTCCGGGGGGG  657

Qy    601 GGCAGG---CTGCTGGAGCAGGAGAGGGCGGCCGAGCTGCCGCCGCCCTTCGTGGTGCTCTGC  657
            ||||||||||||||||||||
Db    689 GGGCAAGCTCCCTGGGGCGGAGAGTGCCGGGCATGCCGCCATCACGCGCCTCACACTGTGGTCT  748

Qy    658 ATCGAGAAACAGCGGTCATGACCTCCTTCTCCCAAG--TAGGGCCGGGGCCCCACCGGACCAGG  715
            ||||||||||||||||
Db    749 ATTGAGAAACAGCTTCATGACTGCCCGTC-CAAGTAGCGCCGCCGCCGGGGGGACCCCG  808

Qy    716 CGGGGGAGGGCGGGGGCCGCCCGAGGAGCCATCCGCCCGCCCCGGGGGGGCCTGG  773
            ||||||||||||||||||
Db    809 AAAAGGCCCCGCAACCTCCGAAGAAACCCCCACCCTGGGGCCCCGGAATTGGCCCC  866
```

Query Match   54.4%; Score 451; DB 13; Length 979;
Best Local Similarity 78.8%; Pred. No. 1.1e-55;

```
TITLE        Sequence evaluation of four pooled-tissue normalized bovine cDNA
             libraries and construction of a gene index for cattle
JOURNAL      Genome Res. 11 (4), 626-630 (2001)
MEDLINE      21180013
PUBMED       11282978
COMMENT      Contact: Smith TPL
             USDA, ARS, US Meat Animal Research Center
             PO Box 166, Clay Center, NE 68933-0166, USA
             Tel: 402 762 4356
             Fax: 402 762 4390
             Email: smith@email.marc.usda.gov
             Single pass sequencing. Bases called and alt trimmed with phred
             v0.980904.e. Vector identified by cross_match with the -minscore 18
             and -minmatch 12 options.
             PCR Primers
             FORWARD: AGGAAACAGCTATGACCAT
             BACKWARD: GTTTTCCCAGTCACGACG
             Plate: 81   row: F  column: 9
             Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..551
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 2BOV"
                     /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
ORIGIN

Query Match         53.7%;   Score 445.2; DB 10;   Length 551;
Best Local Similarity   88.5%;  Pred. No. 7.2e-55;
Matches 483;  Conservative  0;  Mismatches  63;  Indels  0;  Gaps  0;

Qy    163 GTGGTGCACCTGGTTGGCCCTGAACAGCCCGCAGCCGGGCCATGCGAGGCATCCGGGGA 222
          |||||||||||||||||||||| ||||||||||||||||||||||| |||||||||||||
Db      1 GTGGTGCACCTGGTGGCGCTCAACAGCCCGCAGTCGGGCCGGGCGGGGCCATCCGCGGC 60

Qy    223 GCGGACTTCCAGTGTTCCAGCAGGCGGCGCGCCGGGGCTGCGCCGGCACCTTCCGGGCC 282
          ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db     61 GCCGACTTTCAGTGCTTCCAGCAGGCGGCGCGGGGCTGCGGCGGCACCTTCCGGGCG 120

Qy    283 TTCCTGCTGCGCGGGCTGCCAGGACCTTCTACAGACATCGTGCGCCGGCCGCCACCGGG 342
          ||||| |||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    121 TTCCTGTCCTGCCGGTTGCGAGGACCTTGTACAGACATCGTGCGCCGGGCCGGGTGCCACC 180

Qy    343 GTGCCGTGCGTCAACCTCAGGGACGAGAGGTGCTCTTCCCCCAGCTGGGAGGCCTATTCTCG 402
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    181 CTGCCCGTGCGTCAACCTCAGGGACGAGAGGTGCTTTTCCTAGCTGGAGGCCCTTGTTCTCA 240

Qy    403 GGCTCCGAGGGCCAGCTGAAGCCCGGGGCGCCCGGGGCCCGCATCTTCTCTTTCGACGGCAGAGATGTC 462
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGCTCCGAGGGCCAGCTGAAGCCCGGGGCGCCCGGGGCTCCGCATCTTCTCTTTCGACGGCAGAGATGTC 300

Qy    463 CTGAGCACCCCGCCTGGCCCGGGAAAGAGCGTGTGGCACGGCTCCAGCCCCAGCGGGCGC 522
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTTCAGCATCCCACCTGGCCCCAGAAGAGCGTGTGGCACGGCTCATACCCCAGCGGGGCGC 360

Qy    523 CGGCTGACCGACAGCACTACTGCGAGACGTGGCGACGAGCCCGGGCGGCCACCGGGCAG 582
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CGGGTGACCGAGAGCTACTGCGAGACGTGGCGGGACGAGAGGCGGCCCACCAGGCCAG 420

Qy    583 GGCGTCGTCGCTGCTGGGCGGCGCAGGCTGCTGGAGCAGGAGGCGGAGCTGCGGCCACGCC 642
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GCCTCCTCGCTGCTGGGCGGCCCGTGCTGCTGGAGCAGAAAGCCCGGGGTGCCACAACGCC 480

Qy    643 TTCGTGGTGCTCTGCATCGAGAACAGCTTCATCATGACTTCCTTCCAAGTAGGAGGGCCCGCGCG 702
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTCATTGTCCTCTGCATCGAGAACAGCTTCATGACCTTCCTCCCAAGTAGGAGGGCCCTCTGC 540
```

```
Qy    703 GCCCAC 708
          |||||
Db    541 GGGCAC 546

RESULT 6
BG387051
LOCUS        BG387051                   832 bp    mRNA      linear   EST 12-MAR-2001
DEFINITION   602454749F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4582933 5',
             mRNA sequence.
ACCESSION    BG387051
VERSION      BG387051.1  GI:13280497
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 832)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM1306  row: a   column: 14
             High quality sequence stop: 679.
FEATURES             Location/Qualifiers
     source          1..832
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4582933"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_15"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
ORIGIN

Query Match         53.3%;   Score 441.6; DB 12;   Length 832;
Best Local Similarity   82.1%;  Pred. No. 2.5e-54;
Matches 550;  Conservative  0;  Mismatches 104;  Indels 16;  Gaps  3;

Qy      1 CCCTGCGGGCAGATGACATCCTGGTCCGCCCCCGCCGGCTGTGGACCCCAGCCCTAC 60
          ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db     89 CCCTGGCGGGCAGATGACATCCTGGCCCAGCCCCGCCGGCTGCCCGAGCCCCAGCCCCTAC 148

Qy     61 CCCGGGGGCCCGGCCGGACACCACGGCGCCTCCTAGTGACTTCCTACGTGCACCCACTGGTGGG 120
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    149 CCCGGAGCCCCGGACACCACGCTCCTACGTGCACCTCCTACGTGCACCACCACAAG---- 203

Qy    121 CCCGTCCACACCCACACCACCACCACCAGGACTTCAGCTGCTGCACCTGGTGGCC 180
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    204 -------CCACACCGCCCACAGCACCCGGACTTCCAGCCGTGCTGCTCCACCTGGTTGCG 255

Qy    181 CTGAACAGCCCGCAGCATGCAGGCATCGGGGAGCGGACTTCCAGTGCTTC 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 CTCAACAGCCCCCTGTCAGGCGGCGCATGCGGGGCATCGGGGCCGACTTTCCAGTGCTTC 315

Qy    241 CAGCAGGCGGCCCGCGCGCGGGGTTGGCCGGCACCTTGGCCGCACCTTCCGTGTCGTCGGCGCTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAGCAGGCGGCCCGCGCGCGGGGTTGGCCGGCACCTTGGCCGCACCTTCCGTGTCGTCGGCGCTG 300
```

```
Db   316 CAGCAGGCGCGGGGCCGTGGGGCTGCGGGCACCTTCCGCGTCCTCGGCCTG   375
Qy   301 CAGGACCTCTACAGCATCGTGCGCGGCGCGGGGCTGCGGGTGTCAACCTC     360
         ||||||||||||||||| ||| || | |||  |||||||||||||||
Db   376 CAGGACCTGTACAGCATCGTGCGCCATCGTGCGCCGTGCCGCCGGGCTGCGGGTGTCAACCTC   435
Qy   361 AGGGACGAGGTGCTCTTCCCCAGTGGGAGGGCCTTATTCTCGGGCTCGAGGGCCAGCTG   420
         ||||||||| ||||| |||||| ||||||| |||  | ||  || ||  || ||| |||
Db   436 AAGGACGAGCGTGCTGTTTCCCAGCTGGGAGGCTCGTGTTCTCAGGGCTCTGAGGGTCCGCTG   495
Qy   421 AAGCCCCGGGCCCCGCATCTTCTTTTCGACGGGCAGAGATGTCCTGACGGCACCCCCGCCTGG   480
         ||||||||| |  |||||||||| |||  |||||| ||  | ||| ||  ||||||| ||||
Db   496 AAGCCCCGGGCACCGCATCTTCTCTCCTTTGACGGCCAAGGACGTCCTGAGGCACCCCACCTGG   555
Qy   481 CCCCGGAAGAGAGCCGTGTGGCGCACGGCTCCGACCCCCAGCGGGCGGCC--GCCTGACCGACCAGCT   538
         ||| | ||||||||| ||||||| | |||  |||||||| |  ||||||| ||||||| | || |||
Db   556 CCCCAGAAGAGAGCGCGTGTGGGCATGGCGTCGGACCCCCAACCGGGCGGCCGCAGGCTGACCCCGGAGAGCT   615
Qy   539 ACTGCGAGACGTGGCGGACGGAGGCCCCGGGCCGCCACCGGGGCGAGGCGTGTGCGTGCTGG   598
         |||| ||||||||||||||| ||| || |||   |||  ||||| | | ||  ||| |||||
Db   616 ACTGTGAGACGTGGCGGACGGAGGCTCCCTGGCACACGGCCCCAGGGCCTC-TCGCTGCTGG   674
Qy   599 CGGGCAGGCTGCTCGGAGCAGGAGGCCGCCGGAGCTGCCGCCAGGCCTTCGTGGTGGTTCGCTCTGCA   658
         | ||||||  | | |||| ||||| || |||||||||| || ||   |||  ||||| || ||||||
Db   675 GGGGCAGGGTTCCTCGGGGGCAGGAGTGCCGAGAGTGTGCCCTGCCTCACGCTTACATCGTGTCTCTGCA   734
Qy   659 TCGAGAACAG   668
         | ||||||||
Db   735 TTGAGAACAG   744
```

```
RESULT 7
BC062931

LOCUS       BC062931                4192 bp    mRNA    linear   HTC 17-DEC-2003
DEFINITION  Mus musculus cDNA clone IMAGE:30354418, containing frame-shift
            errors.
ACCESSION   BC062931
VERSION     BC062931.1  GI:38614356
KEYWORDS    HTC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4192)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 4192)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-NOV-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
```

```
REMARK
COMMENT     NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            cDNA Library Preparation: M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
            Thomas L. Casavant.
            Web site: http://genome.uiowa.edu
            Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
            Bonaldo,M.F., Akabogu,I., Bair,T., Bair,J., Crouch,K., Davis,A.,
            Fishler,K., Keppel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K.,
            Scheetz,T., Smith,C., Snir,E., Tack,D., Trout,K., Walters,J.,
            Casavant,T., Soares,M.B.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: Plate:   Row:   Column: 0
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 13937350
            This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1..4192
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30354418"
                     /tissue_type="Brain, mouse, 13.5,14.5,16.5,17.5 dpc"
                     /clone_lib="NIH_BMAP_PY0"
                     /lab_host="DH10B"
                     /note="Vector: pYX-ASC"
ORIGIN

Query Match             52.7%;  Score 436.8;  DB 11;  Length 4192;
Best Local Similarity   73.5%;  Pred. No. 1.5e-53;
Matches  625;  Conservative   0;  Mismatches 192;  Indels  33;  Gaps   4;

Qy     1 CCCTGGCGGGCAGATGACATCCTGGCCCGCCCCCCGGCGGCCTGCTGGACCCCCAGCCCTAC     60
         ||||||||| || ||||| ||||| ||| | |||  ||||||| |||| |||||||||||||
Db  3313 CCCTGGCGGAGCAGAGTGACAGCATCCTGGCCAACCACCACCACCGGCGCGGCTGGCCTCGGCCAGCGCAGCCTTAC   3372
Qy    61 CCCGGGG--CCCCGGCACCAGGCTCCTACGTGCACTCCATGTGCACCTGCCGGCCCAGGCCCGGGCCCGCCCAC---    117
         |||||||  |||| | |||| || || | ||||||  | || ||||||| ||||||||| ||| || ||||||
Db  3373 CCCGGAGTTCCACATCACCACACAGTTCCTATGTGCACCTGCACCTGCCGGCCCAGCTGGTGCTGGCTCACCTGGT   3428
Qy   118 GGGCCCGTCCACACCCCCACACCCACCACCCACCACCAGGACTTCCAGCTGGTGGCTCACCTGGTG    177
                     ||||| | | | || ||| | | |||||||||||||||| ||||| ||||||||
Db  3429 -------CCTCTCACTTGCTCATACTCATCAGGACTTTCAGCTGGTGCTCCACCTGGTG    3480
Qy   178 GCCCTGAACAGCCCGGCGCCGGCCGATGGAAGGCATCGGGAGGAGAGCGGCATCCGGGCGGCGGACTTCCAGTGC    237
         || |||||||| ||   | |||| || ||| || |||| |||||||| || |||  |||| ||||| ||||||||
Db  3481 GCACTGAACAACCCCCCTGTCTGGAGGCATGGCGTGGTATCCGTGGAGCAGATTTCCAGTGC    3540
Qy   238 TTCCAGCAGGGCGGCGGCCCGGGGGCTGGCCGGCGGCGCACCTTCGGGCGGGCCTTCCGTGTCGTCGGGG    297
         ||||||||  ||||  |||||||||||||| ||||||||  |||||||||| ||||| ||||| | |
Db  3541 TTCCAGCAAGCCGGAGCCCGGGGGCTGGGGCGTGTCGGGGCTGTGGGCCACCTTCGGGGCGCTTTTCGTGTCCTGTAGG   3600
Qy   298 CTGCAGGACCTTCACAGACATCGTGCGCGGCCGCACCGGGTGCCGGTCGTCAAC    357
         |||||||| || | | ||||||||||| |||  | ||||| || | |||||||
Db  3601 CTGCAGGATCTCTATAGACATCGTGCGGCCGCTGCTGACCGGGGGTGTCTGTGCCGCCATCGTCAAC   3660
Qy   358 CTCAGGACGACGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTCGGGCTCGAGGGCCAG    417
         || |||||| ||||||| ||| |||||||||||||| || |||  | ||||| ||| ||||
Db  3661 CTGAAGGACGAGGTGCTATCTCCCCAGCTGGGAGGACTCCGTCTTTCTGGCTCCGGGGTCAA   3720
Qy   418 CTGAAGCCCCGGGGCCCGCATCTTCTTTGACGGCAGAGATGTCCTGCAGGCACCCCGCC    477
         ||| || ||||||||||||||| || ||||||  | ||||||| || |||||| |||
Db  3721 CTGCAACCCCGGGGCCCGCATCTTTTCTTTGACGGCAGGCAGAGATGTCCTGAGACACCCAGCC   3780
Qy   478 TGGGCCCCGGAAGAGACGGTGTGGCCACGGCGTCCGACCCCCAGCGCGCGGGCGCCCTGACCGACACCAGCAG    537
         ||||||  ||||||| | | ||||  ||||| ||||||||||||  |||| | | |||  | | | ||| |||
Db  3781 TGGGCCGCAGAAGAGAGCGTATGGCACGGCTCGGACCCCCAGTGGGCGGGAGGCTGATGGAGAGT   3840
```

```
Qy    538 TACTGCGAGACGTGGCGACGAGGCCCCGGCGGGCCACCGGCGGGCGCGGTGCGTGCGGTGCGTGCTG 597
          ||||||   ||||||  || |||||||||||    |||||  |||||||||||||
Db   3841 TACTGTGAGACATGGCGGAACTGAAACTACTGGGGGCTACAGGTCAGGGCTGGGGGCTCCCCTGCTG 3900

Qy    598 GCGGGCAGGCTGCTGGAGCAGGAGGAGGCCCGGAGCTGCCGCCACGCCCTTCGTCGTGCTCTGC 657
          |||| ||||||  | || ||||||  ||||    ||  || |||||| ||||||  ||||
Db   3901 TCAGGCAGGCTCCTGGAACAGAAAGCTGCGACTGCGACTGCCACACAGCTACATCGTCCTGTGC 3960

Qy    658 ATCGAGAACAGCGTCATGCGACTCCTTCTTCCAAGTAGGGCCGGCGGCCACGGACAGGCG 717
          || ||||||||||   ||||||   |||||||||||||   ||||||  || |||||||
Db   3961 ATTGAGAATAGCTTCATGACCTCTTTCTTCCCAAATAGGGCCTCTGCCCAGCTAGGGTGGCAG 4020

Qy    718 GGGGAGG-------------GGGGCCCCGCAGGAGCAGCATCCGCGCCCCGCGGGGGGCC 761
          |||||||             ||||| |||||||||||||||||||||||||||
Db   4021 ACAGAGGGCCATGCAGCAGAACTTTGACACAGCGCAGGGAGCATTCAGTCAGTCAGCACCCAGGGCTC 4080

Qy    762 TGGCCGGGACGCTTGCCTGCCACCGT--CACGTTTAATGTAATCCTCAAGAATAAAAAGGA 819
          |||||||||||   |||||      ||  ||  ||||   ||| |||||||||| |||||||
Db   4081 TGGCTGGGATACAATCCTGTATAGTTCCCATTTTTATGTAATCCTCAAGAATAAAAAGGA 4140

Qy    820 AGCCAAAGAG 829
          ||||||||||
Db   4141 AGCCAAAGAG 4150
```

RESULT 8
BF385854

LOCUS   BF385854   843 bp   mRNA   linear   EST 27-NOV-2000
DEFINITION 602046021F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4195660 5',
           mRNA sequence.
ACCESSION   BF385854
VERSION   BF385854.1 GI:11367159
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 843)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Jeffrey E. Green, M.D.
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM9531 row: i column: 05
           High quality sequence stop: 761.
FEATURES   Location/Qualifiers
     source   1..843
                 /organism="Mus musculus"
                 /mol_type="mRNA"
                 /strain="FVB/N"
                 /db_xref="taxon:10090"
                 /clone="IMAGE:4195660"
                 /clone_lib="NCI_CGAP_Li9"
                 /lab_host="DH10B (TI phage-resistant)"
                 /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                 Average insert size 1.9 kb. Constructed by Life
                 Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

Query Match   51.6%; Score 428; DB 10; Length 843;
Best Local Similarity 77.9%; Pred No. 2.3e-52;
Matches 547; Conservative 0; Mismatches 140; Indels 15; Gaps 2;

```
Qy      1 CCCTGCGGGGGCGAGATGACATCCTGGCCGGCCGCCCCCGGCCGGCTGCCGACCCCCAGCCCTAC 60
```

```
Qy     42 CCCTGGCGAGCAGCAGATGACATCCTGGCCAACCCACCCACCGCGCCCGCTTGCCGCCAGCCTTAC 101
          ||||||||   |||||||||||||||||||| ||||
Db     42 CCCTGGCGAGCAGCAGATGACATCCTGGCCAACCCACCCACCGCGCCCGCTTGCCGCCAGCCTTAC 101

Qy     61 CCCGGG---CCCCGCACCACGGCTCCTACGTGCACTTCAGCGGCTTGCCGCCCACTGGT 117
          |||||||   || ||||||  | ||||||  |||||||  ||||| ||| |||||||
Db     61 CCCGGG---CCCCGCACCACGGCTCCTACGTGCACTTCAGCGGCTTGCCGCCCACTGGT 117

Qy    102 CCCGGAGTTCCACATCACCACAGTTCCTATGTGCACCCGCCCCAGCCCGCCCAC---- 157
          |||||||  |||||  ||| |||||| |||||||||||||||||||||||||
Db    102 CCCGGAGTTCCACATCACCACAGTTCCTATGTGCACCCGCCCCAGCCCGCCCAC---- 157

Qy    118 GGGCCCGTCCACACCCCACACCCACCCACCAGGACTTCCACTGGTGCTGACACCTGTG 177
          |||||||| |||| ||||||||| |||||| |||| ||||||   |||  |||||||
Db    118 GGGCCCGTCCACACCCCACACCCACCCAGGACTTCCACTGGTGCTGACACCTGTG 177

Qy    158 -------CCTCTCACTTGCTCATATCTCATCAGGACTTTCAGCCAGACTTTCCACCTGTG 209
                 ||||| ||| | || || |||| ||| |||||||| |||||||| ||||||||
Db    158 -------CCTCTCACTTGCTCATATCTCATCAGGACTTTCAGCCAGACTTTCCACCTGTG 209

Qy    178 GCCCTGAACAGCCCGCAGCCGGGGGCATGCGGGGCATCCGGGGAGCCGGGGAGACTTCCAGTGC 237
          |  |||||||    |  ||      |||||||||||||  ||||||||  |||||||||||||||
Db    178 GCCCTGAACAGCCCGCAGCCGGGGGCATGCGGGGCATCCGGGGAGCCGGGGAGACTTCCAGTGC 237

Qy    210 GCACTGAACACCCCCCTGTCTGGAGGCATGCGTGGCGTGGGAGACAGATTTCCAGTGC 269
          || |||||||||||||||||||||||||||||| |||||||||||| |||||||||||
Db    210 GCACTGAACACCCCCCTGTCTGGAGGCATGCGTGGCGTGGGAGACAGATTTCCAGTGC 269

Qy    238 TTCCAGCAGGCGCGCGCGGGCGGCTGCCCGGCGACCTTCCGGGCCTTCCGTCGTGCGCGG 297
          ||||||||  |||||||||  |||||||||||| |||||||||||||||||||| |||||
Db    238 TTCCAGCAGCCCGACCCCGAGCCCGGCTGCGTGTGGGCACCTTCCGGGCACCTTCCTCTCTAGG 329

Qy    298 CTGCAGGACCTTACAGCATCGTGCGCGCCAGCTGGGAGGCCACCGGGGGTCTGTGCCCGTCGTCAAC 357
          |||| |||  |  | ||||||||  |||||||||||||||    |||  | ||  ||||||||||||
Db    298 CTGCAGGACCTTACAGCATCGTGCGCGCCAGCTGGGAGGCCACCGGGGGTCTGTGCCCGTCGTCAAC 357

Qy    330 CTGCAGGATCTCTATAGACATCGTGCCGCCCGTGCCTGACCGGGGGTCTGTGCCCATCGTCAAC 389
          ||||||||  ||  | ||||||||  |  | ||  ||||| |||  | ||  ||||||||||||
Db    330 CTGCAGGATCTCTATAGACATCGTGCCGCCCGTGCCTGACCGGGGGTCTGTGCCCATCGTCAAC 389

Qy    358 CTCAGGGACAGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTTCGGGCTCCGAGGGGCCAG 417
          |||||||   |  || |||||||||||| |||||||||   ||||||||||||| |||||||||
Db    358 CTCAGGGACAGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTTCGGGCTCCGAGGGGCCAG 417

Qy    390 CTGAAGGACGAGAGTGCTATCTCCCAGCTGGACTCCCAGCTTTTCTGGCTCCCCAGGGGTCAA 449
          ||  ||||  ||| ||| ||||  ||||| |||| |||||  |||||||||  |||||||||||
Db    390 CTGAAGGACGAGAGTGCTATCTCCCAGCTGGACTCCCAGCTTTTCTGGCTCCCCAGGGGTCAA 449

Qy    418 CTGAAGCCCGGGCCGCCGCATCTTCTCTTTCGACGGCCAGAGATGTCCTGCAGCACCCCGCC 477
          |||  |||||||  | ||  ||| |  ||| | |||||||||||| |||||||||||||||||
Db    418 CTGAAGCCCGGGCCGCCGCATCTTCTCTTTCGACGGCCAGAGATGTCCTGCAGCACCCCGCC 477

Qy    450 CTGCAACCCCGGGCCCGCCATCTTTTTCTTTTGACGGCCAGAGATGTCCTGAGACACCCAGCC 509
          ||||| |||| |||| |||||| ||  ||| | |||||||||||||||||| | |||||||||
Db    450 CTGCAACCCCGGGCCCGCCATCTTTTTCTTTTGACGGCCAGAGATGTCCTGAGACACCCAGCC 509

Qy    478 TGGCCCCGGAAGAGAGCGTGTGGCACGCGGCCTCCGACGGCGCCCGCTGACCGGCCGACCAGAGC 537
          ||| |||||||||||||| |||||||| ||||| |||| || |||  ||| |||||| ||||||
Db    478 TGGCCCCGGAAGAGAGCGTGTGGCACGCGGCCTCCGACGGCGCCCGCTGACCGGCCGACCAGAGC 537

Qy    510 TGGCCGCAGAAGAGCGTATGGCACGGCTCGGAACCCCAGTGGGCGGAGGCGAGTGATGGGAGAGT 569
          ||||| |||||||||| |||||||| ||||||| || | ||| || |||| ||||| ||||||||
Db    510 TGGCCGCAGAAGAGCGTATGGCACGGCTCGGAACCCCAGTGGGCGGAGGCGAGTGATGGGAGAGT 569

Qy    538 TACTGCGAGACGTGGCGGACGAGGCCCCGGCGGGCCACCGGCGGGCGGTGCGTGCGTGCTG 597
          |||||||| |||||||  |||||||  ||||| ||  || |||||||   || || ||||||
Db    538 TACTGCGAGACGTGGCGGACGAGGCCCCGGCGGGCCACCGGCGGGCGGTGCGTGCGTGCTG 597

Qy    570 TACTGTGAGACATGCGGACTGAAACTACTGGGGCTACAGGCTACAGGTCAGGGCTCCCCTGCTG 629
          |||||||||||||  |||||||||||| ||| |||||||| ||||||||||||||  ||||||||
Db    570 TACTGTGAGACATGCGGACTGAAACTACTGGGGCTACAGGCTACAGGTCAGGGCTCCCCTGCTG 629

Qy    598 GCGGGCAGGCTGCTGGAGCAGGAGGAGGCCCGAGCTGCCGCCACGCCCTTCGTCGTGGTCTGC 657
          || || ||||| ||||| |||||  | ||||| ||||  || ||||| || | |||  ||||||
Db    598 GCGGGCAGGCTGCTGGAGCAGGAGGAGGCCCGAGCTGCCGCCACGCCCTTCGTCGTGGTCTGC 657

Qy    630 TCAGGCAGGCTCCTGGAACAGAAAGCTGCGACTGCCACAACAGCTACATCGTCCTCTGTGC 689
          ||||| |||||| |||||| |||| || | ||  |||||||||||||| |||||  |||||
Db    630 TCAGGCAGGCTCCTGGAACAGAAAGCTGCGACTGCCACAACAGCTACATCGTCCTCTGTGC 689

Qy    658 ATCGAGAACAGCGTCATGCGACTCCTTCTTCCAAGTAGGGCCGC 699
          ||||||||  ||  |||| | || || ||||||| |||||| ||
Db    690 ATTGAGAATAGCTTCATGACCTCTTTCTTCCCAAATAGGGCCTC 731
```

RESULT 9
AU125614

LOCUS   AU125614   715 bp   mRNA   linear   EST 01-AUG-2002
DEFINITION AU125614 NT2RM4 Homo sapiens cDNA clone NT2RM4001897 5', mRNA
           sequence.
ACCESSION   AU125614
VERSION   AU125614.1 GI:10950330
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 715)
  AUTHORS   Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
           Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
           Isogai,T.
  TITLE   HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
           Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T.,
           Suzuki,Y., Sugano,S., Isogai,T.)
  JOURNAL   Unpublished (2000)
COMMENT   Contact: Takao Isogai
           Genomics Laboratory
           Helix Research Institute
           1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
           Tel: 81-438-52-3975

Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES   Location/Qualifiers
     source   1..715
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RM4001897"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /clone_lib="NT2RM4"
                     /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
                     precursor cells"

ORIGIN

Query Match   51.4%;   Score 425.8; DB 9; Length 715;
Best Local Similarity 82.3%; Pred. No. 4.7e-52;
Matches 506; Conservative 0; Mismatches 97; Indels 12; Gaps 1;

```
Qy    1  CCCTGCGGGCAGATGACATCCTGGCCGCGCCCCCGGCGCGCTGCTGACCCCCAGCCCTAC   60
Db   99  CCCTGCGGGCAGATGACATCCTGGCCAGCCCCCTCCCGCTGCCCCGAGCCCCAGCCCTAC  158

Qy   61  CCCGGGGCCCCCGCCACCACGGCTCCTACGTGCACTTCCAGCCGGCTCGCCCCACTGGTGGG  120
Db  159  CCCGGAGCCCCCGCACCACAGCTCCTACGTGCACTGCGCGGCCGGCACGACCCACAG----  214

Qy  121  CCCGTCCACACCCACACCCACCACCCACCAGGACTTCCAGCTGGTGCTGCGACCTGGTGGCC  180
Db  215  --------CCACCCGCCCACAGCCACCGGACTTCCAGCCGGTTCCACCTGCTTGCCACCCGGTTGCG  266

Qy  181  CTGAACAGCCCGCCAGCCGGCGGCGGCATGCGGAGGCATCCGGGAGCGGGACTTCCAGTCTTC  240
Db  267  CTCAACAGCCCCTGTCAGGCGGCATGCGGGGCATCCGCGGGGCCGACTTCCAGTGCTTC  326

Qy  241  CAGCAGGCGGCGGCCGCGGCGGGGCTGGCCGGGCACCTTCCGGGGCCTTCCGTGCGTGCGGGCTG  300
Db  327  CAGCAGGCGGCGGCCGGTGGGCTGGCGGGCACCTTCCGGGGCACCTTCCGTGTCCGTGCTCGTCAACCTC  386

Qy  301  CAGGACCTTCTACAGGATCGTGCGGCCACGGCTCCGACCGGTCCCGTCGTCAACCTC  360
Db  387  CAGGACCTGTACAGCATCGTGCGGCCGTGCCGACCGCGCAGCCCGTGCCCATCGTCAACCTC  446

Qy  361  AGGGACGAGGTGCTTCTCCCCAGCTGGGAGCCCTTATTCTCGGGCTTCGGAGGGCCAGCTG  420
Db  447  AAGGACGAGCTGCTGTTTCCCAGCTGGGAGAGGCTCTGTTTCTCAGGCTCGGAGGGTCCGCTG  506

Qy  421  AAGCCCGGGGCCCCCATCTTCTTTTCGACGGCAGAGATGTCCTGCAGCACCCGCCCTGG  480
Db  507  AAGCCCCGGGGCACGCACCATCTTTCTCTTTTGACGGCAAGGACGTCCTGAGGCACCCCACCTGG  566

Qy  481  CCCCCGGAAGAGCGTGTGGCACGGCTCCGACCCCAGCGGGGCGCCGGCGCTGACCGCAGCTAC  540
Db  567  CCCCAGAAGAGCCGTGTGGCATGGCTCGGACGTCGGAGGCTCGACCGAGGCTAC  626

Qy  541  TGCGACGTGGCGGACGGAGGCCCCGGGGCCGCACCGGGCCAGGAGGCGTCGTCGCTGCTGGGCG  600
Db  627  TGTGAGACGTGGCGGACGGAGGCCCCGGGGCCCCCTCCCTTCCGGCACGGGGCACGGCCTTCTCCGGTTGCTGGGGG  686

Qy  601  GGCAGGCTGCTGGAG  615
Db  687  GGCAAGCTTCTTGGG  701
```

ACCESSION   BU556872
VERSION   BU556872.1 GI:22907168
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 947)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/;
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2795 row: f column: 04
            High quality sequence stop: 616.

FEATURES   Location/Qualifiers
     source   1..947
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6584644"
                     /tissue_type="teratocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_109"
                     /note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
                     XhoI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAGG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC library."

ORIGIN

Query Match   51.4%;   Score 425.8; DB 13; Length 947;
Best Local Similarity 84.2%; Pred. No. 4.9e-52;
Matches 508; Conservative 0; Mismatches 82; Indels 13; Gaps 2;

```
Qy    1  CCCTGCGGGCAGATGACATCCTGGCCGCGCCCCCGGCGCGCTGCTGACCCCCAGCCCTAC   60
Db  136  CCCTGGCGGGCAGATGACATCCTGGCCCAGCCCCCTCCCGCTGCCCCGAGCCCCAGCCCTAC  195

Qy   61  CCCGGGGCCCCCGCCACCACGGCTCCTCCTACGTGCACTTCCAGCCGGCTCGCCCCACTGGTGGG  120
Db  196  CCCGGAGCCCCCGCACCACAGCTCCTACGTGCACTGCGCGGCCGGCACGACCCACAAG----  251

Qy  121  CCCGTCCACACCCACACACCCACCACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTTGGCC  180
Db  252  -------CCCACCCGCCCACAGCCACCGCGACTTCCAGCCGGTTCCACCTGCTGTCACCTGGTTGCG  303

Qy  181  CTGAACAGCCCGCCAGCCGGCGGCATGCGGAGGCATCCGGGAGCGGGACTTCCAGTCGTTC  240
Db  304  CTCAACAGCCCCTGTCAGGGCGCATGCGGGGCATCCGCGGGGCCGACTTCCAGTGCTTC  363

Qy  241  CAGCAGGCGGCGGCCGCGGCGGGGCTGGCCGGGCACCTTCCGGGGCCTTCCGTCGTCGTGCGGCTG  300
Db  364  CAGCAGGCGGCGGCGGCCGGTGGGCTGGCGGGCACCTTCCGGGGCACCTTCCGTGTCCGTGTCGTCGGCTG  423

Qy  301  CAGGACCTCTACAGGATCGTGCGGCCGACCGGTCCCGGCCGTGCCCGTCGTCACCTC  360
Db  424  CAGGACCTGTACAGCATCGTGCGGCCGGCCGACCGCGCAGCCCGTGCCCATCGTCAACCTC  483

Qy  361  AGGGACGAGGTGCTTCTCCCCAGCTGGGAGGCCCTTATTCTCGGGCTCCGAGGGCCAGCTG  420
Db  484  AAGGACGAGCTGCTGTTTCCCAGCTGGGAGGCTTCAGGCTCGTGTTCTCAGGCCTCAGGCCTCTGAGGGGTCCGCTG  543
```

RESULT 10
BU556872
LOCUS   BU556872   947 bp   mRNA   linear   EST 16-SEP-2002
DEFINITION AGENCOURT_10186760 NIH_MGC_109 Homo sapiens cDNA clone
            IMAGE:6584644 5', mRNA sequence.

RESULT 11

```
BE908201
LOCUS       BE908201                 707 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601500458F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3902175 5',
            mRNA sequence.
ACCESSION   BE908201
VERSION     BE908201.1  GI:10402537
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 707)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9704 row: d  column: 16
            High quality sequence stop: 688.
FEATURES             Location/Qualifiers
     source          1..707
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3902175"
                     /tissue_type="epithelioid carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_70"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.1 kb. Library constructed by Life
                     Technologies."
ORIGIN
```

Query Match   51.0%; Score 422.8; DB 10; Length 707;
Best Local Similarity   78.2%; Pred. No. 1.3e-51;
Matches 508; Conservative   0; Mismatches 142; Indels   0; Gaps   0;

```
Qy    129 CACCCACACCCACACCCACCAGGACTTCCAGCTGGTGCTCACCTGGTGGCCCTGAACCAG 188
          ||| |||  ||| |||||  ||||||||||||||||||||||||||||| ||| ||||||
Db      1 CCCACCCGCCCACAGCCGCCCACGCCGACTTCCAGCTGGTGCTCACCTGGTTGCGCTCACCAG 60

Qy    189 CCCGCAGCCGGCGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGGC 248
          |||  |||||||||  ||||  |  ||||||  ||||||||||||||||||||||||||
Db     61 CCCCCTCTCAGCGGCGGCATCCCGGGGCCATCCCGCGGGCCGACTTCCAGTGCTTCCAGCAGGC 120

Qy    249 GCGGCGCGGGGGCTGGCCGGCCGCACCTTCCGGTCGTCGTGCGCGGCTGCAGGACCT 308
          ||| |||||||| |||||  |||||||||||| |||| ||| | ||  ||||||||||
Db                                                                  

Qy    421 AAGCCCGGGGCCGGCCGCCGCCATCTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCGCCTGG 480
          |||||||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    544 AAGCCCGGGGCACGGCGACGATCTCTCCTTGCGAGGCAAGGACGTCCTTGAGGCACCCACCTGG 603

Qy    481 CCCCGGAAGAGCGGTGGCACGGCGGCCCCGGCCCCGGGCGCGGGCGGGC 540
          ||||  |  |||||||||||||||||||||||  |  ||||||||||||
Db    604 CCCCAGAAGAGCGTGTGGCATGGCTCGGACCCCAACCCGGGCGCAGGCTGACCGAGAGCTAC 663

Qy    541 TCGGAGACGTGGCGGACGAGGCCCCGGCGGCCACCGGACCGGCGGTCGTCGTGCTGCTGGC 599
          ||||||||||||||||||||||  |||| |||||||||  |||||||||||||||||||||
Db    664 TGTGAGACGTGGCGGACGCGGAGGCTCCTCGGCCACCGCAGGAGGGCCTCCTCGGCTGTGGG 723

Qy    600 GGG 602
          |||
Db    724 GGG 726
```

RESULT 12

```
BE906253
LOCUS       BE906253                 757 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601502237F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3904208 5',
            mRNA sequence.
ACCESSION   BE906253
VERSION     BE906253.1  GI:10399595
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 757)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9709 row: i  column: 09
            High quality sequence stop: 757.
FEATURES             Location/Qualifiers
     source          1..757
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3904208"
                     /tissue_type="epithelioid carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_70"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
```

```
Qy    121 GCGGGCCGTGGGGGTCGGCGGGGCGCACCTTCCGCGCTTCCTGTCCTCGGCCGCTGCAGGACCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    309 CTACAGCATCGTGGCGCGCCGACGCCACGGGGTGCCCGTGCTCGTCAACCTCAGGGACGA 368

Qy    181 GCGGGCCGTGGGGGTCGGCGGGGCGCACCTTCCGCGCTTCCTGTCCTCGGCCGCTGCAGGACCT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    369 GGTGCTTCTTCCCCAGCTGGTAGGCCTTATTCTCCGGGTTCCGAGGGCCAGCTGAAGCCCGG 428

Qy    241 GGTGCTTCTTCCCCAGCTGGTAGGCCTTATTCTCCGGGTTCCGAGGGCCAGCTGAAGCCCGG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    429 GGCCCGCATCTTCTTTCGACGGCAGAGATGTCTCGCAGCACCCGCCTGGCCCCGGGA 488

Qy    301 GGCCCGCATCTTCTTTCGACGGCAGAGATGTCTCGCAGCACCCGCCTGGCCCCGGGA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    489 GAGCGTGTGGCACGGCTCCGACCCCAACCCGGGCGCAGGCGGCGTGTCTCGCTCGGGGGCAGGCT 548

Qy    361 GAGGCGTGTGGCATGGCTCGGACCCCCAACGGGCGCAGGGGCTGACCGAGAGCTACTCTGAGAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    549 GTGGGGGACCGAGGCCCCGGCGCACCGGGCAGGCGTCGTCGCTCGGTGCTCTGCATCGAACAG 608

Qy    421 GTGGGGGACCGAGGCGCCCGGCGCACGGGCCAGGCCTCCGCTGGTGGTGCTCTCGCATCGAACAG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609 GCTGGAGCAGGAGGCGCGGAGCTGTCGCGCCCACGCCTTCGTCGGTGCTCTGCATCGAACAG 668

Qy    481 GCTGCTGTTTCCCACGTGGCATGGCTCGGAGGCTCTGTTCTCAGGTCGTCTGCTGCTGGGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    669 CCTGGGCCAGAGTGCCGGAGTGCCCATCACCGCCTACACATCGTGCTCTGCATTGAGAAACAG 728

Qy    541 CGTCATGACCTCCTTCTCCAAGTAGGGCCGCGGCCCACCGGACACAGGCGGGGGAGGGGGGC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    729 CTTCATGACTGCTCCAAGTAGCCACCGCTGGATGCAGATCGCCGGAGAGATCGGCGACGGTTGCC 778

Qy    601 GCCCCGAGGAGCATCCGCCGCCCCGGGGGGCCTGGCCGGACGGTTGCC 650
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTCGGAGGAAGCCCCAACGTGGGCCAGGGAGCAGCCGCCGGCCAGCCCCTTGGC 650
```

Average insert size 1.1 kb. Library constructed by Life
Technologies."

Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by:Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11361 row: g column: 18
High quality sequence start: 23
High quality sequence stop: 808.

FEATURES   Location/Qualifiers
source   1..874
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="Czech II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5146409"
                     /tissue_type="pooled lung tumors"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NCI_CGAP_Lu33"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site 1: NotI; Site 2: EcoRI; 1st
                     strand cDNA was prepared from mRNA obtained from pooled
                     lung tumors with a Not I - oligo(dT) primer (5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTCTGTTTTTTTTTTTTTT 3').
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match   Score 417; DB 12; Length 874;
Best Local Similarity 50.3%; Pred. No. 9e-51;
Matches 537; Conservative 0; Mismatches 140; Indels 16; Gaps 2;

Qy   11 CAGATGACATCTGGCCGGCCCCCCCGCCGCCTGCTGGACCCCAGCCCTACCCCGGGG--- 67
          |||||||||||||||||| |||||||| |||||||||||||||| |||||||||||
Db   23 CAGATGACATCTGGCCAACCCCACCCGGCGCTGCCAGACCGCCGCCTTACCCCGGGAGTT 82

Qy   68 -CCCCGCAACCACGGCTCCTACGTGCACTTCCAGCCGGCTCGCCCACTGGTGTGGGCCCGTC 126
           ||||||| |||||| |||| || ||||||||||||||| |||| |||| |
Db   83 CCACATCACCACAGTTCCTATGTGCACTTCGCCCGCCAGCCGGCCCCAC----------C 130

Qy   127 CACACCCACACCCACCCACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCCCTGAAC 186
          |||||||||||||||||||| ||||||||||||||| |||||| |||||||||||| ||||
Db   131 CTCTCACTGGTCTCATACTCATCAGGACTTTCAGCCAGTTCAGCCAGTCTCCACCTGGTGGCACTGAAC 190

Qy   187 AGCCCGCAGCCGGCGGCGGCATGGAGGCATCCGGGAGGCATCCGGGGAGCGGACTTCAGGCAG 246
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   191 ACCCCCCGTCTGAGGCATGCGGTGGTATCCGTGGAGCAGAGATTTCAGTGCTTCCAGCAA 250

Qy   247 GCGGCGGCCGCGGCCGTGGCGGGCTGGCCGGACGCCTTCCGTGCGTCGCCGGCTGCAGGAC 306
          |||||||||||| ||||||||||||| |||||||||||||||||||||||||||||||
Db   251 GCCCGGAGCCGGTGGGGCGGGCTGTCGGGGCACCTTCCGGGGCACCCGGCTTTCCGGCAG 310

Qy   307 CTCTACAGCATCGTGCGCACAGCATCGTGGGAGCCTGGAGGGTGCCCGTCGTCAGGAGG 366
          ||||||| ||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   311 CTCTATAGCATCGTGCGCGCGTCGACCGGGGGTCTGTGCCATCGTCAACCTGAAGGAC 370

Qy   367 GAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTCGGGGCTCCGAGGGCCAGCTGAAGCCC 426
          ||||||||||||||||||||||||| ||| ||||||||||||||||||||||||||||
Db   371 GAGGTGCTATCTCCCAGCTGGGACTGCCCTGTTTCTGGCTCTCCAGGGTCAACTGCAACCC 430

Qy   427 GGGGGCCCGCATCTTCTTCTTTGACGGGCAGAGATGTCTCAGCACCCCGCCTGGCCCCGG 486
          |||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   431 GGGGGCCCGCATCTTTCTTTCTTTGACGGCAGCAGAGATGTCCTGACAGCACCCAGCCTGGCCGCAG 490

Qy   487 AAGAGGCGTGTGCACGGCTCCGACCCCAGCCAGCCGGCGCGCCGCTCGACCGACAGCTACTGCGAG 546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   491 AAGAGCGTATGGCACGGCGTCCGGACCCCGACCCCCAGTAGCGGGCGGAGAGGCTGATGCAGCTGCAACCC 550

Qy   547 ACGTGGCGGACGGAGGAGCCCCCGGGGCACCCGGGCCCAGGGCCACCGGCTCGCTGGTGGCGCGGCAGG 606
          ||||||| ||| ||||| |||| |||||||||||||| || ||||| ||||||||||||||||

ORIGIN

Query Match   Score 422.4; DB 10; Length 757;
Best Local Similarity 51.0%; Pred. No. 1.5e-51;
Matches 504; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy   119 GGGCCGTCCACACCCACACCCACCCACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGG 178
          |||||||||||||||||||||||||| |||||||||||||||||| ||||||||||||||||
Db   16 GACCCACAAGCCACCCGCCCACCCACAGCCCACCGGACTTCCAGCCGGTGCTCCACCTGGTTG 75

Qy   179 CCCTGAACAGCCCGCAGCCGGGCATGCGAGGCATCCGGGGAGGCCGGACTTCCAGTGCT 238
          || || ||||||||| ||| ||||| ||||||||||||||||| |||||||||||||||
Db   76 CGGTCAACAGCCCCCGTCAGGCGGCATGCGGGGCATCCGGGGCCGGGCACTTCCAGTGCT 135

Qy   239 TCCAGCAGGCGGCGGCCGCGGCCGTGGGCGGGCTGGCCGGTGCCACTTCGGGGCCTTCGTCGCCGGC 298
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   136 TCCAGCAGGCGGCGGCCGCGGCCGTGGGGCGGGCACCTTCCGGCGGCCTTCGTCGTCCTGCGCC 195

Qy   299 TGCAGGACCTTCACAGCATCGTGCGCCGCGCGCGACCGCACCGGGGTGCCGTCGTCAACC 358
          ||||||||||| |||||||||||||||||||| ||| || ||||||||||||||||||
Db   196 TGCAGGACCTGTACAGCATCGTGCGCGCGTCGACCGCGCGCGTGCCCATCGTCAACC 255

Qy   359 TCAGGGACGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTCGGGGCTCCGAGGGCCAGC 418
          |||| ||||||||||| |||||||||||||||||| |||||||||||||||||||||||
Db   256 TCAAGGACGAGCTGCTGTTTCCCAGCTGGGAGGCTCTGTTCTCAGGCTCTGAGGGTCCGC 315

Qy   419 TGAAGCCCGGGGGCCCGCATCTTCTTCTTTGACGGACGAGATGTCCTGCAGCACCCCGCCT 478
          ||||||||||||||||||| ||||| ||||||||||||||||||||||||||||||||||
Db   316 TGAAGCCCGGGGCACGCATCTTCTTCTTTGACGGACGAGATGTCCTGAGGCACCACCACCT 375

Qy   479 GGCCCCGGAAGAGGCGTGTGGCACGGCTCCGACCCCAGCGGCGGCCGCCGACCAGT 538
          |||||||||||||||||||| |||||||| ||||||| ||||||||||||||||||
Db   376 GGCCCCAGAGAGAGCGTGTGGCATGGCGTGCGGGACCCCAACGGGCGCCAGCGGGCGCCGAGAGCT 435

Qy   539 ACTGGCGAGACGTGGCGGACCAGAGACCCCCGGCGGCCACCGGCGACGAGGGTCGCGTGCTGG 598
          ||||||| || ||||||||||||| ||||| |||||||||||||||||| |||| |||||||
Db   436 ACTGTGAGACGTGGCGGACCGAGGAGCTCCCTCGCGCCCAGGCCTCCGGCGCTCCGGTGCTGG 495

Qy   599 CGGGCAGGGCTGCTGAGCAGCAGGAGGCCGGAGCGTGCCGCCCACAGCCTTCGTGGTGCTCGCA 658
          ||||||||| |||||||| ||||||||||||||||| ||| ||||||||||||||||||||
Db   496 GGGGCAGGGCTTCCTGGGGCAGAGTGCGCGGAGCTGCGCCATCACGGCTACATCGTGCTCTGCA 555

Qy   659 TCGAGACACGGTCATGGACCTCCTTCTCCAAGTAGGAGGCGGCGCCCACGGACGAGGGCGG 718
          |||| || ||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   556 TTGAGAACAGTTCATGATCGCTTCCAAGTAGCCACCGCCTTGATGCAGATGGCCGGAGA 615

Qy   719 GGGGAGGGGGGGCCCGCAGGAGCAATCCGCCGCCCCCCGGGGGGG 758
          |||||||||||||||||||||||||||||||||||||||||||
Db   616 GGACCCGGCGGCTCGGAGGAAGCCCCCCACCGTGGCAGGGAG 655

RESULT 13
BI412588
BI412588   602990468F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5146409 5',
            mRNA sequence.
LOCUS   BI412588   874 bp   mRNA   linear   EST 14-AUG-2001
DEFINITION 602990468F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5146409 5',
            mRNA sequence.
ACCESSION   BI412588
VERSION   BI412588.1   GI:15173511
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 874)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgaps-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

RESULT 14

```
Db  551 ACATGGCGAACTGAAACTACTGGGGCTACAGGTCAGGGCCTCCTCCCTGCTCGTGTCAGGCAGG 610
        ||||||||||||||||||||||||||||||||||| |||||||||||||||| ||||||| |||
Qy  607 CTGCTGGAGCAGGAGGCCCGCAGAGCTGCCGAGCCTCGTGTGCTGCATCGAGAAC 666
        |||| ||||  |||| ||  || ||||||||| | || |  | |  |||||||| |||
Db  611 CTCCTGGAACAGAAAGCTGCGAGCTGCCACAACAGCTACATCGTCCTGTGCATTGAGAAT 670
Qy  667 ACGGTCATGACCTCCTCTTCCAAGTAGGGCGGC 699
        |||||||||||| ||||||||||| |||||||| ||
Db  671 AGCTTCATGACCTCTTTCTCCAAATAGGGCGGCTC 703
```

```
RESULT 14
CF789984
LOCUS       CF789984            504 bp    mRNA    linear   EST 21-OCT-2003
DEFINITION  867322 MARC 4PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   CF789984
VERSION     CF789984.1  GI:37794545
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1 (bases 1 to 504)
  AUTHORS   Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,
            Nonneman,D.J., Wray,J.E. and Keele,J.W.
  TITLE     Porcine EST collection using a normalized library constructed from
            embryos representing early developmental stages
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified with
            cross_match v0.990329.
            Plate: TMW8006 row: B column: 3
            Seq primer: GTAATACGACTCACTATAGGG.
FEATURES             Location/Qualifiers
     source          1..504
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 4PIG"
                     /note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
                     Library made with combined RNA from day-10, day-13,
                     day-15, day-25, and day-30 whole embryos."
ORIGIN
```

```
Query Match        49.8%;   Score 413;   DB 14;   Length 504;
Best Local Similarity  89.0%;   Pred. No. 3.2e-50;
Matches  446;   Conservative   0;   Mismatches  55;   Indels   0;   Gaps   0;

Qy  200 GGGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGGCGCGCCGCGG 259
        |||| | ||||| ||||||| |||||||||||| ||||||||||||||||| || ||||
Db    1 GCGGGCCTGCGCGGCCATCCGGCGCCGGACTTCCAGCGGCTTCCAGCAGGCGCGCGCCGTGG 60
Qy  260 GGCTGGCGCGGGACCTTCCGGGGCCTTCCGTGCGTGCGGCTGCAGGACCTCTACAGCATCG 319
        ||||||||| || ||||||||| ||||||| | ||| |||||||||||||||||||||||||
Db   61 GGCTGGCGGGGCACCTTCCGGGCCTTCCGTCCGTCCTGCGGGCTGCAGGACCTCTACAGCATCG 120
Qy  320 TGCGCCGGCCCGACCGGCACCGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTCC 379
        ||||  ||| |||||||||||| | | | ||  || |||||||||||||||||||| ||| ||
Db  121 TGCGACGGGCCCGACCGGCACCGCCCGTGCCCATCGTAACCTCAGGGACGAGGTGCTGTTCG 180
Qy  380 CCAGTCGGGAGGCCTTATTCGGGCTCCGAGGGCTCCGAGGGCTGAAGCCCCGGGCCCGATCT 439
        || || ||||| || || || ||||| |||||| |||||| |||| | |||||| ||||||||
Db  181 GGGGCGGGGAGGCGGGCCTTGTTCTCGGGGCTCTGAGGGCCAGCCCAGGGGCCAGCGCCCGATCT 240
Qy  440 TCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCCCTGGCCCCCGGAAGAGAGCGTGTGGC 499
```

RESULT 15

```
Db  241 TCTCTTCCGACGGCCAGAGACGTCCTTCAGCACCCTGGCCGACCCTGGCCCCCAGAAGAGACGGTGTGGC 300
        ||||||| |||||| |||||  |||| | ||||| |||| ||||| ||||||| ||||||| |||||||||
Qy  500 ACGGGCTCCGACCCCAGCGGGCGCGCGGCGCCGGCTCGGACCCTGACCGACAGCTACTGCGGAGCGTGGCCGGACACG 559
        |||||||| ||||||||||||||| || |||||||||||| | ||||| |||||||||||| || ||||| |||||||
Db  301 ACGGGCTCAGACCCCAGCGGGCGCGGGCGCCGGCTCGGACCGGAGCTACTGTGAGACGTGGCGGGACTG 360
Qy  560 AGGGCCCGGCGGGCACCCGGCAGGCGTCGTCGGTGCTGCTCGGCGGCAGGCGTGCTGAGCAGG 619
        |||| || || |||| |||||||||||||| || |||| ||||||||| || ||| |||||||
Db  361 AGGCCAGGACGGCCACGGGCCAGGCCTCCTCGCGTCGCCGCTGCGGCGGCCGGCGCTGCTGGAGCAAA 420
Qy  620 AGGGCCGGAGCTGCGCCACGCCCGCCCTTCGTCGTGTGCTGCTGCTCGTCGCATCGAGAACAGCGTCATGACCT 679
        ||||||||  || ||||||  ||||||||| | | ||| ||| | |  ||||||||||||||||  |||||||||
Db  421 AGGGCCGGCCGGCTGCCACCACGCCCTTCATTGTTCTCTGCATCGAGAACAGCTTCATGACCT 480
Qy  680 CCTTCTCCAAGTAGGGCCCGG 700
        || ||||||||||||| | ||
Db  481 CCGCCCTTCCAAGTAGGATCTCG 501
```

```
RESULT 15
CB444165/c
LOCUS       CB444165            657 bp    mRNA    linear   EST 25-MAR-2003
DEFINITION  695295 MARC 6BOV Bos taurus cDNA 3', mRNA sequence.
ACCESSION   CB444165
VERSION     CB444165.1  GI:29233914
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 657)
  AUTHORS   Smith,T.P.L., Roberts,A.J., Echternkamp,S.E., Chitko-McKown,C.G.,
            Wray,J.E. and Keele,J.W.
  TITLE     A second set of bovine ESTs from pooled-tissue normalized libraries
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified with
            cross_match v0.990329.
            Plate: FQY8051 row: I column: 11
            Seq primer: TAGAAGGCACAGTCGAGG.
FEATURES             Location/Qualifiers
     source          1..657
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 6BOV"
                     /note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
                     Library made with RNA pooled from multiple tissues
                     including liver, lung, hypothalamus, pituitary, and
                     placenta/endometrium."
ORIGIN
```

```
Query Match        48.4%;   Score 401.6;   DB 14;   Length 657;
Best Local Similarity  81.5%;   Pred. No. 1.5e-48;
Matches  464;   Conservative   0;   Mismatches  105;   Indels   0;   Gaps   0;

Qy  218 GGGGAGCGGACTTCCAGTGCTTCCAGCAGGCGCGCCGCGGCTGGCGCGGCTGGCCACCTTCC 277
        |||| |||| ||| |||||||||||||||||||| || ||||  || ||||  || ||| |||
Db  657 GCGGGCCGCCGACTTTCAGTGCTTCCAGCAGGCGCGGCGCCGGCATCGTCGCGCGGCCGACCATTCC 598
Qy  278 GGGGCCTTCCTGTCGTCGCGCTGCAGGACCTCTACAGCATCGTCGCGCCGGCGCCGACCCGA 337
        || | |||| ||| ||| || |||||| |||||||||||||||||||||||  ||||| |||
Db  597 GCGCGTTCCTTGCGGTTGCAGCAGAACCTCTACAGCATCGTCGCGCCGGCCGCCGTG 538
```

```
Qy  338  CCGGGGTGCCCGTCGTCAACCTCAGGGACAGAGGTGCTCTCCCCAGCTGGGAGGCCTTAT  397
              ||  ||||||||||||||||||||||| |||||||| ||  || |||||||||||||
Db  537  CCACCCTGCCCGTCGTCAACCTCAGGGACGAGGTGCTGTTTCCTAGCTGGGAGGCCTTGT  478

Qy  398  TCTCGGCGCTCCGAGGGCCAGCTGAAGCCCGGGCCCGCATCTTCTTTCGACGGCAGAG    457
              ||||| |||||||||||||||||||||| ||| ||||||||||| |||||||||||||
Db  477  TCTCAGGCTCCGAGGGCCAGCTGAAGCCCGGCGCCCGCATCTTCTTCCTTCGACGGCCAGAG  418

Qy  458  ATGTCCTGCAGCACCCCGCCTTGGCCCCGGAAGAGCGTGTTGCACGGCTCCGACCCCAGCG  517
              |||||| |||||| ||| ||| |||| || ||||||||||| |||| ||| |||||||||
Db  417  ATGTCCTTCAGCATCCCACCTGGCCCCAGAAGAGCGTGTGCACCGCTCAGACCCCAGCG    358

Qy  518  GGCGCCCCTGACCGACAGCTACTGCCGAGACGTGGCGGACGGAGGCCCCGGCGGCCACCG   577
              |||||| |||||||| ||||||||||||||||||||||| || ||| || ||||||| ||
Db  357  GGCGCCGGCTGACCGAGAGCTACTGCCGAGACGTGGCGGACGCAGCAGGCGGCCACGG    298

Qy  578  GGCAGGCCGTCGTCGCTGCTGGCGGGCAGGCTGCTGGAGCAGGAGGCCGCGAGCTGCCGCC   637
              ||||||| | ||||||||||||||||| ||||||||||||| ||  || ||| ||||||||
Db  297  GCCAGGCCTCCTCGCTGCTGGCGGGCGGCTGCTGGAGCAGCAGAAAGCCGCGGGGTGCCACA   238

Qy  638  ACGCCTTCGTGTGTGCTCTGCATCGAGAACACGTCATGACCTCCTTCTCCAAGTAGGGCC   697
              |||||||| | ||| | |||||||||||||| || ||||||||||| ||||||||||||||
Db  237  ACGCCTTCATTGTCCTNTGCATCGAGAACAGCTTCATGACCTCCTCCTCCAAGTAGGGCC    178

Qy  698  GCCGGGCCCACGGACACAGCGGGGGAGGGGGCATCCGCCCGGGGG                  757
              |||  |||  |  |||||| |||||||| |||
Db  177  TCTGCGGGCACCATGGACAGGTGGACTGGTGAACAGAGTGCGAGAGAGAGGCCAGGCGCC   118

Qy  758  GGCCTGGCCGGGACGCTTGCCTGCACCGT  786
              |||   | || ||    |||  ||  ||
Db  117  AGGCAGGAGGAGCGGAATGCCCCCAGCCT   89


Search completed: March 29, 2004, 14:52:44
Job time : 2570.3 secs
```

OM protein - protein search, using sw model

Run on: March 26, 2004, 13:28:41 ; Search time 57.2222 Seconds
                                      (without alignments)
                                      1135.676 Million cell updates/sec

Title: US-09-938-391-2
Perfect score: 1239
Sequence: 1 PWRADDILAGPPRLLDPQPY.........CRHAFVVLCIENSVMTSFSK 230

Scoring table: BLOSUM62
                  Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match   Minimum Match 0%
                 Maximum Match   Maximum Match 100%
                 Listing first 45 summaries

Database :   A_Geneseq.29Jan04:*
              1: geneseqp19808:*
              2: geneseqp1990s:*
              3: geneseqp2000s:*
              4: geneseqp2001s:*
              5: geneseqp2002s:*
              6: geneseqp2003as:*
              7: geneseqp2003bs:*
              8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	1239	100.0	230	5	AAO17429	Aao17429 Canine pr
2	1019	82.2	275	5	AAU76689	Aau76689 Synthetic
3	1019	82.2	310	5	AAU76688	Aau76688 Human col
4	1019	82.2	682	5	ABP41878	Abp41878 Human ova
5	1019	82.2	1301	2	AAW92296	Aaw92296 Human alp
6	1019	82.2	1336	2	AAY08694	Aay08694 Human col
7	1019	82.2	1336	6	ABP96308	Abp96308 Human end
8	1019	82.2	1516	5	ABB83471	Abb83471 Human col
9	1019	82.2	1516	5	ABP68617	Abp68617 Human pan
10	1003	81.0	684	2	AAW26327	Aaw26327 Human alp
11	1003	81.0	684	5	AAO17357	Aao17357 Human col
12	982	79.3	684	5	AAY25113	Aay25113 Human alp
13	980.5	79.1	1288	2	AAW26328	Aaw26328 Mouse alp
14	976	78.8	1288	2	AAW92297	Aaw92297 Mouse alp
15	966	78.0	184	5	AAO17430	Aao17430 Canine en
16	959	77.4	184	3	AAY70265	Aay70265 Canine an
17	940	75.9	684	3	AAY25114	Aay25114 Mouse alp
18	869	70.1	195	3	AAW90874	Aaw90874 Human HMW
19	868	70.1	193	3	AAW90877	Aaw90877 Human HMW
20	850	68.6	184	3	ABG31794	Abg31794 Murine en
21	843	68.0	191	3	AAB28398	Aab28398 Murine en
22	843	68.0	191	5	AAU77950	Aau77950 Amino aci
23	838.5	67.7	271	2	AAB08407	Aab08407 A human a
24	836	67.5	216	3	AAB30495	Aab30495 Amino aci
25	835	67.4	181	4	AAU00898	Aau00898 Human End
26	835	67.4	182	3	AAY59622	Aay59622 Human end
27	835	67.4	182	3	AAY94323	Aay94323 Human end
28	835	67.4	182	3	AAB28399	Aab28399 Human end
29	835	67.4	182	4	AAU00897	Aau00897 Human End
30	835	67.4	182	5	AAU77951	Aau77951 Amino aci
31	835	67.4	183	2	AAY02113	Aay02113 SEQ ID 76
32	835	67.4	183	3	AAY08693	Aay08693 Human end
33	835	67.4	183	3	AAY70252	Aay70252 Human ang
34	835	67.4	183	3	AAY90771	Aay90771 Human ang
35	835	67.4	183	3	AAB16451	Aab16451 Human end
36	835	67.4	183	3	AAB30493	Aab30493 Amino aci
37	835	67.4	183	4	AAB49379	Aab49379 Human end
38	835	67.4	183	4	AAU00896	Aau00896 Human End
39	835	67.4	183	5	ABB79901	Abb79901 Human end
40	835	67.4	183	5	AAM49503	Aam49503 Human end
41	835	67.4	183	5	AAM48895	Aam48895 Human end
42	835	67.4	183	5	AAU97132	Aau97132 Human end
43	835	67.4	183	6	AAG79753	Aag79753 Human End
44	835	67.4	513	5	ABG73586	Abg73586 Human End
45	834	67.3	180	4	AAU00899	Aau00899 Human End

                              ALIGNMENTS

```
RESULT 1
AAO17429
ID   AAO17429 standard; protein; 230 AA.
XX
AC   AAO17429;
XX
DT   19-JUL-2002  (first entry)
XX
DE   Canine pro-endostatin.
XX
KW   Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;
KW   psoriasis; rheumatoid arthritis; retinopathy; macular degeneration;
KW   corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
KW   rubeosis; Osler-Webber Syndrome; myocardial angiogenesis;
KW   plaque neovascularisation; telangiectasia; haemophiliac joints;
KW   angiofibroma; wound granulation; coronary collateral;
KW   cerebral collateral; arteriovenous malformation;
KW   ischaemic limb angiogenesis; diabetic neovascularisation; fracture;
KW   cytostatic; antiarthritic; antipsoriatic; antidiabetic; ophthalmological;
KW   gynaecological.
XX
OS   Canis familiaris.
XX
PN   EP1191036-A2.
XX
PD   27-MAR-2002.
XX
PF   24-AUG-2001; 2001EP-00307224.
XX
PR   25-AUG-2000; 2000US-0227924P.
XX
PA   (PFIZ ) PFIZER PROD INC.
XX
PI   Sheppard MG, Tong X;
XX
DR   WPI; 2002-354068/39.
DR   N-PSDB; AAL46062.
XX
PT   An isolated nucleic acid molecule for the treatment of angiogenesis-
PT   related disorder, such as cancers or diabetic retinopathy, encodes an
PT   endostatin protein.
XX
PS   Claim 14; Fig 3; 56pp; English.
XX
CC   The present invention provides the protein and coding sequences of canine
CC   pro-endostatin and endostatin. The sequences can be used in the treatment
CC   and diagnosis of angiogenesis related disorders, including cancer,
CC   rheumatoid arthritis, psoriasis, retinopathy, macular degeneration,
```

```
CC   corneal graft rejection, neovascular glaucoma, retrolental fibroplasia,
CC   rubeosis, Osler-Webber Syndrome, myocardial angiogenesis, plaque
CC   neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,
CC   wound granulation, coronary collaterals, cerebral collaterals,
CC   arteriovenous malformations, ischaemic limb angiogenesis, diabetic
CC   neovascularisation, and fractures. The present sequence is the canine pro
CC   -endostatin protein sequence
XX
SQ   Sequence 230 AA;

Query Match      100.0%;    Score 1239;  DB 5;   Length 230;
Best Local Similarity   100.0%;   Pred. No. 3.2e-131;
Matches  230;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 PWRADDILAGPPRLLDPQPYPGAPHHGSVHFQPARPTGGPVHTHTHQDFQLVLHLVA 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 PWRADDILAGPPRLLDPQPYPGAPHHGSVHFQPARPTGGPVHTHTHQDFQLVLHLVA 60

Qy    61 LNSPQPGGWRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 LNSPQPGGWRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120

Qy   121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180

Qy   181 CETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 230
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 230

RESULT 2
AAU76689
ID   AAU76689 standard; protein; 275 AA.
XX
AC   AAU76689;
XX
DT   21-MAY-2002 (first entry)
XX
DE   Synthetic plasmid pEnd-HR#1 FPD fusion protein sequence.
XX
KW   Mouse; Ig signal peptide; mIgSP; functional protein domain; FPD;
KW   primary translational product; PTP; DNA construct; regulatory DNA;
KW   DNA targeting segment; regulatory factor; single regulatory unit;
KW   monoclonal antibody; recombination-derived alteration; blood product;
KW   human; COL18A1; mutant; mutein; fusion protein.
XX
OS   Mus sp.
OS   Homo sapiens.
OS   Synthetic.
OS   Chimeric.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /label= Signal peptide
FT                   /note= "Mouse Ig signal peptide (mIgSP)"
FT   Region          1..19
FT                   /note= "Encoded by mouse Ig signal peptide (mIgSP) exon"
FT   Protein         20..275
FT                   /note= "Mature human COL18A1 protein"
FT                   /note= "Contains exons 38-41"
XX
PN   WO200210372-A1.
XX
PD   07-FEB-2002.
XX
PF   01-AUG-2001; 2001WO-GB003455.
XX
PR   01-AUG-2000; 2000GB-00018876.
XX
PA   (ISTF ) ARS APPLIED RES SYSTEMS HOLDING NV.
PA   (CHAP/) CHAPMAN P W.
XX
```

```
PI   Chapman PW, De Luca G, Falciola L;
XX
DR   WPI; 2002-195963/25.
DR   N-PSDB; ABK09977.
XX
PT   Producing functional protein domain by growing host cell transfected with
PT   DNA construct having regulatory DNA and DNA targeting segment, and
PT   optionally culturing homologously recombinant cell and collecting
PT   protein.
XX
PS   Example; Fig 8; 116pp; English.
XX
CC   The present invention relates to a new method of producing a protein,
CC   such as functional protein domain, that is either C- or N-terminus of the
CC   primary translational product (PTP) of a gene, where the protein has
CC   biological activity which is distinct from PTP. The method of the
CC   invention involves growing a host cell transfected with a DNA construct
CC   comprising a regulatory DNA and a DNA targeting segment. This method is
CC   useful for producing a functional protein domain of proteins such as
CC   regulatory factors, blood products and monoclonal antibodies. The method
CC   described in the invention allows controlled and precise modification of
CC   the host cell genome in order to produce functional protein domain (FPD).
CC   The amount of exogenous sequence to be integrated in the host cell genome
CC   is very limited since, as coding sequence, the original coding sequence
CC   present in the host cell genome itself is used. Use of the host cell
CC   sequence encoding FPD also provides the advantages of both eliminating
CC   any recombination-derived alteration of such coding sequence, and also
CC   making use of the same post-transcriptional (e.g., splicing) and/or post-
CC   translational (e.g. glycosylation, phosphorylation) processes that are
CC   actually applied in vivo for the maturation of FPD. The use of a single
CC   regulatory unit eliminates the necessity of manipulating the
CC   complementary DNA coding for the PTP to isolate the segment coding for
CC   the FPD, and adapt it to the expression vector. The present amino acid
CC   sequence represents the plasmid pEnd-HR#1 FPD fusion protein of the
CC   invention. This fusion protein contains the mouse Ig signal peptide
CC   (mIgSP) sequence fused to exons 38-41 of the human COL18A1 sequence
XX
SQ   Sequence 275 AA;

Query Match       82.2%;   Score 1019;   DB 5;   Length 275;
Best Local Similarity   82.4%;   Pred. No. 3.1e-106;
Matches  187;  Conservative  16;  Mismatches  20;  Indels  4;  Gaps  1;

Qy     1 PWRADDILAGPPRLLDPQPYPGAPHHGSVHFQPARPTGGPVHTHTHQDFQLVLHLVA 60
         ||||||||| |||| |||||||||||| || |||||| |||         |:||||||
Db    51 PWRADDILASPPRLPEPQPYPGAPHHSSVVHLRPARPTSPP---AHSHRDFQPVLHLVA 106

Qy    61 LNSPQPGGWRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
         |||| ||||||||||||||||||||| |||||||||||||||||||||||||   |:|||
Db   107 LNSPLSGGWRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL 166

Qy   121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180
         :||||:||||||||||||| :|||||||||||| |:|||:||:||||||||||:||||||:|
Db   167 KDELFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTESY 226

Qy   181 CETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTS 227
         |||||||| :||||||||| ||||          |||:||||| ||:
Db   227 CETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSPMTA 273

RESULT 3
AAU76688
ID   AAU76688 standard; protein; 310 AA.
XX
AC   AAU76688;
XX
DT   21-MAY-2002 (first entry)
XX
DE   Human collagen XVIII 1alpha NC1 domain protein sequence.
XX
KW   Human; collagen XVIII 1alpha NC1 domain; functional protein domain; FPD;
KW   primary translational product; PTP; DNA construct; regulatory DNA;
KW   DNA targeting segment; regulatory factor; single regulatory unit;
```

Best Local Similarity 82.4%; Pred. No. 3.7e-106;
Matches 187; Conservative 16; Mismatches 20; Indels 4; Gaps 1;

```
Qy     1 PWRADDILAGPPRLLDPQPYPGAPHHGSVHFQPARPTGGPVHTHTHQDFQLVLHLVA 60
         ||||||||||.||||| :|||||||||||||:||||   :|:|:|||||||||||||
Db    86 PWRADDILASPPRLPEPQPYPGAPHHSSVHLRPARPTSP2---AHSRDFQPVLHLVA 141

Qy    61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
         |||| ||||||||||||:|| ||||:||||| |||||||||| |||||.|:||| |:|||
Db   142 LNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL 201

Qy   121 RDEVLFPSWEALFSGSEGQLKPGARIFSFPDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180
         :||:||||||||||||||| |:||||||:||||:|: .||||:|:|||||||||||||:|
Db   202 KDELLFPSWEALFSGSEGPLKPGARIFSFPDGKDVLRHPTWPQKSVWHGSDPNGRRLTESY 261

Qy   181 CETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTS 227
         ||||||||.|||||||||||||||.:||: ||||||||||||||||
Db   262 CETWRTEAFSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 308
```

RESULT 4
ABP41878
ID   ABP41878   standard; protein; 682 AA.
XX
AC   ABP41878;
XX
DT   22-AUG-2002   (first entry)
XX
DE   Human ovarian antigen HEEBK29, SEQ ID NO:3010.
XX
KW   Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW   ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW   infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW   PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW   inflammatory condition; immune disorder; blood disorder;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disorder; urinary system disorder; drug screening;
KW   gene therapy; chromosome mapping; forensic analysis;
KW   antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW   antiinflammatory; gynaecological; reproductive; chromosome 21q22.3.
XX
OS   Homo sapiens.
XX
PN   WO200200677-A1.
XX
PD   03-JAN-2002.
XX
PF   07-JUN-2001; 2001WO-US018569.
XX
PR   07-JUN-2000; 2000US-0209467P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE, Rosen CA;
XX
DR   WPI; 2002-147878/19.
DR   N-PSDB; ABQ54955.
XX
PT   Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT   useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
PT   cancer), immune disorders, cardiovascular disorders and neurological
PT   diseases.
XX
PS   Claim 11; SEQ ID NO 3010; 2922pp; English.
XX
CC   The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC   ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC   encompasses polypeptides 90% identical and polynucleotides 95% identical
CC   to the sequences of the invention. The invention additionally relates to
CC   recombinant vectors and host cells comprising human ovarian antigen
CC   polynucleotides, antibodies against human ovarian antigens, and the use
CC   of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC   treating, prognosing or preventing various ovary and/or breast-related

---

KW   monoclonal antibody; recombination-derived alteration; blood product.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   Domain   1..54
FT   /label= Multimerisation_domain
FT   Region   1..43
FT   /note= "Encoded by exon 36"
FT   Region   44..54
FT   /note= "Encoded by exon 37"
FT   Region   55..136
FT   /label= Hinge region
FT   /note= "Encoded by exon 38"
FT   Domain   137..310
FT   /label= Endostatin core domain
FT   /note= "Autonomous_folding unit"
FT   Region   137..201
FT   /note= "Encoded by exon 39"
FT   Region   202..241
FT   /note= "Encoded by exon 40"
FT   Region   242..310
FT   /note= "Encoded by exon 41"
XX
PN   WO200210372-A1.
XX
PD   07-FEB-2002.
XX
PF   01-AUG-2001; 2001WO-GB003455.
XX
PR   01-AUG-2000; 2000GB-00018876.
XX
PA   (ISTF) ARS APPLIED RES SYSTEMS HOLDING NV.
PA   (CHAP/) CHAPMAN P W.
XX
PI   Chapman PW, De Luca G, Falciola L;
XX
DR   WPI; 2002-195963/25.
XX
PT   Producing functional protein domain by growing host cell transfected with
PT   DNA construct having regulatory DNA and DNA targeting segment, and
PT   optionally culturing homologously recombinant cell and collecting
PT   protein.
XX
PS   Example; Fig 3; 116pp; English.
XX
CC   The present invention relates to a new method of producing a protein,
CC   such as functional protein domain, that is either C- or N-terminus of the
CC   primary translational product (PTP) of a gene, where the protein has
CC   biological activity which is distinct from PTP. The method of the
CC   invention involves growing a host cell transfected with a DNA construct
CC   comprising a regulatory DNA and a DNA targeting segment. This method is
CC   useful for producing a functional protein domain of proteins such as
CC   regulatory factors, blood products and monocloral antibodies. The method
CC   described in the invention allows controled and precise modification of
CC   the host cell genome in order to produce functional protein domain (FPD).
CC   The amount of exogenous sequence to be integrated in the host cell genome
CC   is very limited since, as coding sequence, the original coding sequence
CC   present in the host cell genome itself is used. Use of the host cell
CC   sequence encoding FPD also provides the advantages of both eliminating
CC   any recombination-derived alteration of such coding sequence, and also
CC   making use of the same post-transcriptional (e.g. splicing) and/or post-
CC   translational (e.g. glycosylation, phosphorylation) processes that are
CC   actually applied in vivo for the maturation of FPD. The use of a single
CC   regulatory unit eliminates the necessity of manipulating the
CC   complementary DNA coding for the PTP to isolate the segment coding for
CC   the FPD, and adapt it to the expression vector. The present amino acid
CC   sequence represents the human collagen XVIII alpha NC1, a functional
CC   protein domain used to illustrate the method of the invention
XX
SQ   Sequence 310 AA;

Query Match   82.2%;   Score 1019;   DB 5;   Length 310;

```
CC    disorders. Such conditions include ovarian cancer and breast cancer, and
CC    metastatic tumours of ovarian or breast origin, reproductive system
CC    disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC    polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC    disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC    shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC    vaginitis), immune disorders (e.g., congenital and acquired
CC    immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC    blood-related disorders (e.g., anaemia, cardiovascular disorders,
CC    respiratory disorders, neurological disorders, gastrointestinal disorders
CC    and urinary system disorders. Ovarian antigen polypeptides and
CC    polynucleotides may also be used in screening for compounds which
CC    modulate ovarian antigen expression or activity. The polynucleotides may
CC    further be used for gene therapy, chromosome mapping, in the
CC    identification of individuals and in forensic analysis, and the
CC    polypeptides may be used as food additives or to prepare antibodies
CC    useful in disease diagnosis, drug targeting and phenotyping. The present
CC    sequence represents a human ovarian antigen of the invention. Note: The
CC    sequence data for this patent did not form part of the printed
CC    specification, but was obtained in electronic format directly from WIPO
CC    at ftp.wipo.int/pub/published_pct_sequences
XX
SQ    Sequence 682 AA;

Query Match     82.2%;  Score 1019;  DB 5;  Length 682;
Best Local Similarity  82.4%;  Pred. No. 1.1e-105;
Matches  187;  Conservative  16;  Mismatches  20;  Indels  4;  Gaps  1;

Qy    1 PWRADDILAGPPRLLDPQPYPGAPHHGSYHFQPARPTGGPVHTHTHTHQDFQLVLHLVA 60
          |||||||| ||| |||  :|:||| |||||| |||||    :|:|| ||||||
Db  458 PWRADDILASPPRLPEPQPYPGAPHHSSYVHLRPARPTSPP----AHSHRDFQPVLHLVA 513

Qy   61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
          ||||| |||||||||||||||||||:  ||||||||||||||||||||||| |:|| |
Db  514 LNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL 573

Qy  121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGGRRLTDSY 180
          :||:||||||||||| ||:|||||||||||| |::|||  ||| ||||||||  ||:|:
Db  574 KDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTESY 633

Qy  181 CETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTS 227
          |||||||| ||||||||| ||| :::||||||
Db  634 CETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 680


RESULT 5
AAW92296
ID   AAW92296 standard; peptide; 1301 AA.
XX
AC   AAW92296;
XX
DT   28-APR-1999  (first entry)
XX
DE   Human alpha-1 (XVIII) collagen chain common sequence HU18 (common)36.
XX
KW   Human; type XVIII collagen; liver disease; cirrhosis; detection;
KW   hepatocellular carcinoma; diagnosis.
XX
OS   Homo sapiens.
XX
PN   WO9856399-A1.
XX
PD   17-DEC-1998.
XX
PF   12-JUN-1998;  98WO-US012327.
XX
PR   12-JUN-1997;  97US-0049369P.
XX
PA   (FIBR-) FIBROGEN INC.
PA   (FIFI-) ACAD FINLAND.
PA   (INRM ) INST NAT SANTE & RECH MEDICALE.
XX
PI   Pihlajaniemi T,  Rehn M,  Clement B;
```

```
DR   WPI; 1999-070292/06.
XX
PT   Diagnosis and monitoring of liver disease by measuring collagen type
PT   XVIII levels - with elevated levels indicative of disease, especially
PT   cirrhosis or hepatocellular carcinoma.
XX
PS   Example 6; Fig 8; 56pp; English.
XX
CC   A method has been developed for the detecting liver disease. The method
CC   comprises: (a) reacting a patient sample with antibodies (Ab) specific
CC   for collagen type XVIII (Col18); (b) measuring the amount of Ab-antigen
CC   complex (C) formed as indicator of the amount of Col18 present; (c)
CC   similar analysis of a non-diseased control; and (d) comparing the amounts
CC   of Col18 in the two samples to detect presence or progression of disease.
CC   Elevated levels of Col18 are: (i) indicative of disease, specifically
CC   cirrhosis; and (ii) predictive of the prognosis of disease, specifically
CC   hepatocellular carcinoma (there is a relationship between Col18 mRNA
CC   levels and tumour size and necrosis, and survival times are significantly
CC   higher in patients with higher Col18 levels). The method provides non-
CC   invasive, early and accurate diagnosis of liver disease. The present
CC   sequence represents the sequence common to human alpha-1 (XVIII) collagen
CC   chain from the present invention
XX
SQ   Sequence 1301 AA;

Query Match     82.2%;  Score 1019;  DB 2;  Length 1301;
Best Local Similarity  82.4%;  Pred. No. 2.9e-105;
Matches  187;  Conservative  16;  Mismatches  20;  Indels  4;  Gaps  1;

Qy    1 PWRADDILAGPPRLLDPQPYPGAPHHGSYHFQPARPTGGPVHTHTHTHQDFQLVLHLVA 60
          |||||||| ||| |||  :|:||| |||||| |||||    :|:|| ||||||
Db 1077 PWRADDILASPPRLPEPQPYPGAPHHSSYVHLRPARPTSPP----AHSHRDFQPVLHLVA 1132

Qy   61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
          ||||| |||||||||||||||||||:  ||||||||||||||||||||||| |:|| |
Db 1133 LNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL 1192

Qy  121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGGRRLTDSY 180
          :||:||||||||||| ||:|||||||||||| |::|||  ||| ||||||||  ||:|:
Db 1193 KDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTESY 1252

Qy  181 CETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTS 227
          |||||||| ||||||||| ||| :::||||||
Db 1253 CETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 1299


RESULT 6
AAY08694
ID   AAY08694 standard; protein; 1336 AA.
XX
AC   AAY08694;
XX
DT   10-AUG-1999  (first entry)
XX
DE   Human collagen 18 protein.
XX
KW   Plasminogen; human; angiostatin; endostatin; gene therapy; vector;
KW   anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW   tumour growth; solid tumour; diabetic retinopathy; retina; collagen 18.
XX
OS   Homo sapiens.
XX
PN   WO9926480-A1.
XX
PD   03-JUN-1999.
XX
PF   20-NOV-1998;  98WO-US024950.
XX
PR   20-NOV-1997;  97US-00975424.
XX
PA   (GENE-) GENETIX PHARM INC.
PA   (MASI ) MASSACHUSETTS INST TECHNOLOGY.
```

```
XX
PI   Leboulch P, Pawliuk RJ, Bachelot T;
XX
DR   WPI; 1999-357596/30.
DR   N-PSDB; AAX77720.
XX
PT   Anti-angiogenic gene therapy vectors.
XX
PS   Disclosure; Page 77-80; 83pp; English.
XX
CC   This invention describes a novel viral gene therapy vector comprising a
CC   nucleic acid molecule encoding an anti-angiogenic polypeptide chosen from
CC   human or murine angiostatin, human or murine endostatin and angiogenesis-
CC   inhibiting fusions and fragments, where the viral vector is sufficiently
CC   attenuated for use in human gene therapy. The products of the invention
CC   have anti-angiogenic, cytostatic, anti-diabetic and ophthalmological
CC   activity. The vector is used in gene therapy for inhibiting tumour growth
CC   in humans harbouring a solid tumour. The vector expresses an anti-
CC   angiogenic polypeptide. An additional use comprises treatment of diabetic
CC   retinopathy, where the anti-angiogenic polypeptide inhibits angiogenesis
CC   in the vicinity of the retina. The vector is administered to cells ex
CC   vivo and then administered to the patient
XX
SQ   Sequence 1336 AA;

Query Match            82.2%;  Score 1019; DB 2;  Length 1336;
Best Local Similarity  82.4%;  Pred. No. 3e-105;
Matches  187; Conservative  16; Mismatches  20; Indels  4; Gaps  1;

Qy         1 PWRADDILAGPPRLLDPQYPGAPHHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLVA 60
             |||||||| ||||   |  ||||||||| |||| ||||||   : :  ||| ||||||
Db      1112 PWRADDILASPPRLPEPQPYPGAPHHSSYVHLRPARPTSPP---AHSHRDFQPVLHLVA 1167

Qy        61 LNSPQPGGWMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
             |||| ||||||||||||||||||||| || ||||||||||||||||||||||||  :||||||
Db      1168 LNSPLSGGWMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL 1227

Qy       121 RDEVLFPSWEALFSGSBEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180
             :|| |||||||||||||  ||| |||||||||||| |:|| |: ||:|||||||  ||||:|
Db      1228 KDELLFPSWEALFSGSSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTESY 1287

Qy       181 CETWRTEAPAATGQASSLIAGRLLEQEAASCRHAFVVLCIENSVMTS 227
             |||||||| |||||||||  |||| |: ||||: ||::||||||: |:
Db      1288 CETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 1334
```

```
RESULT 7
ABP96308
ID   ABP96308 standard; protein; 1336 AA.
XX
AC   ABP96308;
XX
DT   20-MAY-2003  (first entry)
XX
DE   Human endostatin protein.
XX
KW   Humanised baculovirus; cytostatic; gene therapy; baculovirus; cancer;
KW   prostate cancer; endostatin; chromosome 21.
XX
OS   Homo sapiens.
XX
PN   WO2003016540-A2.
XX
PD   27-FEB-2003.
XX
PF   15-AUG-2002; 2002WO-GB003791.
XX
PR   15-AUG-2001; 2001GB-00019852.
XX
PA   (UYYO-) UNIV YORK.
XX
PI   Maitland N;
XX
```

```
DR   WPI; 2003-268336/26.
XX
PT   New baculovirus having a modified genome encoding a therapeutic agent,
PT   useful in the manufacture of a medicament for the treatment of cancer,
PT   particularly prostate cancer.
XX
PS   Claim 24; Page; 34pp; English.
XX
CC   The present invention describes a humanised baculovirus (I) which
CC   comprises a modified baculovirus genome having a nucleic acid molecule
CC   encoding a therapeutic agent and a polypeptide which functions to target
CC   the baculovirus to at least one cell type. Also described is a
CC   pharmaceutical composition comprising (I). (I) has cytostatic activity,
CC   and can be used in gene therapy. The baculovirus is useful in the
CC   manufacture of a medicament for the treatment of cancer, particularly
CC   prostate cancer. The present sequence represents the human endostatin
CC   protein, which is specified in the exemplification of the present
CC   invention. N.B. The present sequence is not given in the specification
CC   but is referred to in Claim 24 as Genbank accession number NM_130445
XX
SQ   Sequence 1336 AA;

Query Match            82.2%;  Score 1019; DB 6;  Length 1336;
Best Local Similarity  82.4%;  Pred. No. 3e-105;
Matches  187; Conservative  16; Mismatches  20; Indels  4; Gaps  1;

Qy         1 PWRADDILAGPPRLLDPQYPGAPHHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLVA 60
             |||||||| ||||   |  ||||||||| |||| ||||||   : :  ||| ||||||
Db      1112 PWRADDILASPPRLPEPQPYPGAPHHSSYVHLRPARPTSPP---AHSHRDFQPVLHLVA 1167

Qy        61 LNSPQPGGWMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
             |||| ||||||||||||||||||||| || ||||||||||||||||||||||||  :||||||
Db      1168 LNSPLSGGWMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL 1227

Qy       121 RDEVLFPSWEALFSGSBEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180
             :|| |||||||||||||  ||| |||||||||||| |:|| |: ||:|||||||  ||||:|
Db      1228 KDELLFPSWEALFSGSSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTESY 1287

Qy       181 CETWRTEAPAATGQASSLIAGRLLEQEAASCRHAFVVLCIENSVMTS 227
             |||||||| |||||||||  |||| |: ||||: ||::||||||: |:
Db      1288 CETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 1334
```

```
RESULT 8
ABB83471
ID   ABB83471 standard; protein; 1516 AA.
XX
AC   ABB83471;
XX
DT   30-SEP-2002  (first entry)
XX
DE   Human collagen XVIII.
XX
KW   Human; antirheumatic; antiarthritic; gene therapy; anti-angiogenic;
KW   rheumatoid arthritis; collagen; endostatin.
XX
OS   Homo sapiens.
XX
FH   Key              Location/Qualifiers
FT   Protein          1334..1516
FT                    /note= "This region is specifically claimed in Claim 4"
XX
PN   WO200253191-A1.
XX
PD   11-JUL-2002.
XX
PF   03-JAN-2002; 2002WO-KR000001.
XX
PR   05-JAN-2001; 2001KR-00000691.
XX
PA   (VIRO-) VIROMED LTD.
XX
PI   Kim J, Ho S, Park E, Kim S;
```

```
XX
DR   WPI; 2002-583596/62.
DR   N-PSDB; ABN85301.
XX
PT   Novel composition for gene therapy against rheumatoid arthritis,
PT   comprising a DNA encoding anti-angiogenic protein or its parts.
XX
PS   Claim 4; Page 70-78; 84pp; English.
XX
CC   The present invention relates to a composition for gene therapy,
CC   comprising a DNA encoding an anti-angiogenic protein, which shows
CC   therapeutic effects on rheumatoid arthritis. The composition is useful
CC   for treating rheumatoid arthritis and the gene therapy is effective,
CC   lasting for 14 days. The present sequence is the protein sequence for
CC   human collagen XVIII. Endostatin, which consists of the C-terminal 183
CC   residues of collagen XVIII, was used as an anti-angiogenic protein
XX
SQ   Sequence 1516 AA;

Query Match             82.2%;  Score 1019; DB 5; Length 1516;
Best Local Similarity   82.4%;  Pred. No. 3.6e-105;
Matches  187;  Conservative  16; Mismatches  20; Indels  4; Gaps  1;

Qy      1 PWRADDILAGPPRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHTHQDFQLVLHLVA 60
          |||||||||:|||| ||||||||||||:||||  :|:||              :|:||
Db   1292 PWRADDILASPPRLPEPQPYPGAPHHSSYVHLRPARPTSPP----AHSHRDFQPVLHLVA 1347

Qy     61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
          |||| ||||||||||||||||||||:|||||||||||||||||||||||||| :||||||
Db   1348 LNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL 1407

Qy    121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180
          |:||||||||||||||||:|||||||||||||:||||:  :|:||||||||||||||:|
Db   1408 KDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTESY 1467

Qy    181 CETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTS 227
          ||||||||:|||||||||:|||||:  :|||||:|:||||||:||
Db   1468 CETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 1514


RESULT 9
ABP68617
ID   ABP68617 standard; protein; 1516 AA.
XX
AC   ABP68617;
XX
DT   14-JAN-2003  (first entry)
XX
DE   Human pancreatic cancer expressed protein SEQ ID NO 166.
XX
KW   Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;
KW   cytostatic; tumour.
XX
OS   Homo sapiens.
XX
PN   WO200260317-A2.
XX
PD   08-AUG-2002.
XX
PF   30-JAN-2002; 2002WO-US002781.
XX
PR   30-JAN-2001; 2001US-0265305P.
PR   31-JAN-2001; 2001US-0256682P.
PR   09-FEB-2001; 2001US-0267568P.
PR   21-MAR-2001; 2001US-0278651P.
PR   28-APR-2001; 2001US-0287112P.
PR   16-MAY-2001; 2001US-0291631P.
PR   12-JUL-2001; 2001US-0305484P.
PR   20-AUG-2001; 2001US-0313999P.
PR   27-NOV-2001; 2001US-0333626P.
XX
PA   (CORI-) CORIXA CORP.
XX
```

```
PI   Benson DR, Kalos MD, Lodes MJ, Persing DH, Hepler WT, Jiang Y;
XX
DR   WPI; 2002-627435/67.
DR   N-PSDB; ABV94763.
XX
PT   New isolated polynucleotide and pancreatic tumor polypeptides, useful for
PT   diagnosing, preventing and/or treating cancer, particularly pancreatic
PT   cancer.
XX
PS   Claim 2; SEQ ID NO 166; 300pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising: (a)
CC   any of a group of over 4000 nucleotide sequences (ABV94628-ABV99145); (b)
CC   complements of (a); (c) sequences consisting of at least 20 contiguous
CC   residues of (a); (d) sequences that hybridize to (a), under moderately
CC   stringent conditions; (e) sequences having at least 75% or 90% identity
CC   to (a); or (f) degenerate variants of (a) Polypeptides (ABP68596-
CC   ABP68637) encoded by (I) and oligonucleotide can be used to detect cancer
CC   in a patient and compositions comprising polypeptides, polynucleotides,
CC   antibodies, fusion proteins, T cell populations and antigen presenting
CC   cells expressing the polypeptide are useful in treating pancreatic cancer
CC   and stimulating an immune response. The polynucleotides can be used as
CC   probes or primers for nucleic acid hybridisation, in the design and
CC   preparation of ribozyme molecules for inhibiting expression of the tumour
CC   polypeptides and proteins in the tumour cells, in vaccines and for gene
CC   therapy. Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1516 AA;

Query Match             82.2%;  Score 1019; DB 5; Length 1516;
Best Local Similarity   82.4%;  Pred. No. 3.6e-105;
Matches  187;  Conservative  16; Mismatches  20; Indels  4; Gaps  1;

Qy      1 PWRADDILAGPPRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHTHQDFQLVLHLVA 60
          |||||||||:|||| ||||||||||||:||||  :|:||              :|:||
Db   1292 PWRADDILASPPRLPEPQPYPGAPHHSSYVHLRPARPTSPP----AHSHRDFQPVLHLVA 1347

Qy     61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
          |||| ||||||||||||||||||||:|||||||||||||||||||||||||| :||||||
Db   1348 LNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL 1407

Qy    121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180
          |:||||||||||||||||:|||||||||||||:||||:  :|:||||||||||||||:|
Db   1408 KDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTESY 1467

Qy    181 CETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTS 227
          ||||||||:|||||||||:|||||:  :|||||:|:||||||:||
Db   1468 CETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 1514


RESULT 10
AAW26327
ID   AAW26327 standard; protein; 684 AA.
XX
AC   AAW26327;
XX
DT   19-NOV-1997  (first entry)
XX
DE   Human alpha-1 collagen (XVIII).
XX
KW   Alpha-1 collagen; type XVIII collagen; cartilage degeneration.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   Peptide        1..6
FT                  /label= GXYGX'Y' motif
FT                  /note= "Claim 1"
FT   Peptide        7..12
FT                  /label= GXYGX'Y' motif
FT                  /note= "Claim 1"
```

```
FT   Peptide         13..18
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         19..24
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         25..30
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         31..36
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         37..42
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         48..53
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         54..59
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         74..79
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         80..85
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         86..91
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         92..97
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         98..103
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         104..109
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         110..115
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         121..126
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         129..134
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         135..140
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         141..146
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         147..152
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         153..158
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         159..164
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         165..170
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         171..176
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         181..186
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         187..192
FT                   /label= GXYGX'Y'_motif
FT   Peptide         193..198
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         215..220
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         221..226
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         227..232
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         233..238
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         239..244
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         257..262
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         263..268
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         269..274
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         275..280
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         286..291
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         292..297
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         298..303
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         309..314
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         315..320
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         322..328
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         329..334
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         335..340
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         354..359
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         360..365
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         366..372
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         523..528
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         542..547
FT                   /label= GXYGX'Y'_motif
FT                   /note= "Claim 1"
FT   Peptide         590..595
FT                   /label= GXYGX'Y'_motif
```

FT   /note= "Claim 1"
XX
PN   US5643783-A.
XX
PD   01-JUL-1997.
XX
PF   01-DEC-1993;   93US-00159784.
XX
PR   01-DEC-1993;   93US-00159784.
XX
PA   (HARD) HARVARD COLLEGE.
XX
PI   Olsen BR, Oh SP;
XX
DR   WPI; 1997-350247/32.
DR   N-PSDB; AAT84484.
XX
PT   Nucleic acid encoding human alpha-1 collagen - for production of
PT   recombinant alpha-1 collagen, for use in the treatment of cartilage
PT   degeneration.
XX
PS   Claim 1; Col 23-30; 35pp; English.
XX
CC   Novel human type alpha-1 (XVIII) collagen is characterised by 10 triple
CC   helical domains containing the GXGX'Y' motif (where X, Y, X' and Y'
CC   represent any amino acid), the helical domains being separated and
CC   flanked by non-triple helical regions which may provide flexibility.
CC   Alpha-1 collagen is expressed in multiple tissues, especially liver, lung
CC   and kidney. A claimed plasmid comprising alpha-1 collagen nucleic acid
CC   (see AAT84484) and an expression control sequence can be used to express
CC   recombinant collagen in prokaryotic or eukaryotic (especially mammalian)
CC   host cells. The alpha-1 collagen may be used to treat a patient suffering
CC   from a disease associated with cartilage degradation, and for
CC   supplementing collagen. It can also be used as a connective tissue filler
CC   (e.g. for plastic surgery), can be interposed between a dermal equivalent

Query Match   81.0%; Score 1003; DB 2; Length 684;
Best Local Similarity   81.5%; Pred. No. 7.4e-104;
Matches 185; Conservative 15; Mismatches 23; Indels   4; Gaps   1;

Qy   1 PWRADDILAGPPRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLVA 60
           |||||||| ||:|||| ||||||||||| ||||| ||| ||:|| |||||||||
Db   460 PWRADDILASPPGLPBFQPYPGGPHHSSYVHCGPARPTSPP---AHSHRDFQPVLHLVA 515

Qy   61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
           |||| ||||||||||||||| |||||| |||||||||||||||||||| |||||| |||
Db   516 LNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL 575

Qy   121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180
           :||:||||||||| |||| |||||||||| |||| ||| || ||:|||||:|||||:
Db   576 KDELLFPSWEALFSGSEGPLKPGAR1FSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTESY 635

Qy   181 CETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTS 227
           |||||||| |:||||||||| ||| ||:|| |||||||||||||
Db   636 CETWRTEAPSATGQASSLLGGRILGQSAASCHHAYIVLCIENSFMTA 682

RESULT 11
AAO17357
ID   AAO17357 standard; protein; 684 AA.
XX
AC   AAO17357;
XX
DT   19-JUL-2002 (first entry)
XX
DE   Human collagen type XVIII alpha 1.
XX
KW   Human; endometriosis; DNA chip; fibronectin; p27; reticulocalbin;
KW   aldehyde dehydrogenase 6; gravin; phospholipase C epsilon; elastin;
KW   insulin-like growth factor binding protein-2; alpha-2 type IV collagen;
KW   transmembrane receptor PTK7; collagen type XVIII alpha 1;
KW   platelet derived growth factor receptor alpha; laminin M chain;
KW   subtilisin like protein PACE4; nidogen.

XX
OS   Homo sapiens.
XX
PN   EP1191107-A2.
XX
PD   27-MAR-2002.
XX
PF   21-AUG-2001; 2001EP-00250300.
XX
PR   25-SEP-2000; 2000DE-01048633.
XX
PA   (SCHD) SCHERING AG.
XX
PI   Hess-Stumpp H, Haendler B, Kraetzschmar J, Kreft B, Winterhager E;
PI   Regidor P, Scotti S;
XX
DR   WPI; 2002-317413/36.
XX
PT   In vitro diagnosis and monitoring of endometriosis, comprises detecting
PT   reduced expression of specific gene products, e.g. from the fibronectin
PT   gene.
XX
PS   Claim 1; Page 12-13; 21pp; German.
XX
CC   The present invention relates to a method for the in vitro diagnosis of
CC   endometriosis by determining the amount of gene product from at least one
CC   specific gene in a patient sample and comparing this with the amount of
CC   gene product in a control sample. A reduced level is indicative of
CC   endometriosis. The gene products may be fibronectin, p27, reticulocalbin,
CC   aldehyde dehydrogenase 6, gravin, phospholipase C epsilon, elastin,
CC   insulin-like growth factor binding protein-, alpha-2 type IV collagen,
CC   transmembrane receptor PTK7, collagen type XVIII alpha 1, platelet
CC   derived growth factor receptor alpha, laminin M chain, subtilisin like
CC   protein PACE4 or nidogen. The method is useful for initial diagnosis of
CC   endometriosis, and also for monitoring progress and treatment of the
CC   disease. The present sequence is human collagen type XVIII alpha 1
XX
SQ   Sequence 684 AA;

Query Match   81.0%; Score 1003; DB 5; Length 684;
Best Local Similarity   81.5%; Pred. No. 7.4e-104;
Matches 185; Conservative 15; Mismatches 23; Indels   4; Gaps   1;

Qy   1 PWRADDILAGPPRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLVA 60
           |||||||| ||:|||| ||||||||||| ||||| ||| ||:|| |||||||||
Db   460 PWRADDILASPPGLPBFQPYPGGPHHSSYVHCGPARPTSPP---AHSHRDFQPVLHLVA 515

Qy   61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
           |||| ||||||||||||||| |||||| |||||||||||||||||||| |||||| |||
Db   516 LNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL 575

Qy   121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180
           :||:||||||||| |||| |||||||||| |||| ||| || ||:|||||:|||||:
Db   576 KDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTESY 635

Qy   181 CETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTS 227
           |||||||| |:||||||||| ||| ||:|| |||||||||||||
Db   636 CETWRTEAPSATGQASSLLGGRILGQSAASCHHAYIVLCIENSFMTA 682

RESULT 12
AAY25113
ID   AAY25113 standard; protein; 684 AA.
XX
AC   AAY25113;
XX
DT   25-AUG-1999 (first entry)
XX
DE   Human alpha1 (XVIII) collagen protein.
XX
KW   Alpha1 (XVIII) collagen; mimetic; endostatin; atomic coordinate; library;
KW   anti-angiogenic; heparin binding domain; receptor binding domain; mimic;
KW   alpha-helix A domain; carbohydrate recognition domain; CRD domain;

CC This invention describes a novel method for identifying mimetics of
CC mammalian endostatin. The method comprises identifying a compound having
CC atomic coordinates with non-trivial similarity to selected coordinates of
CC atoms of a mammalian endostatin involves (a) providing a library of
CC atomic coordinates of compounds in a library of candidate compounds, (b)
CC comparing the library of atomic coordinates to the selected coordinates
CC of a mammalian endostatin and (c) selecting from the library at least one
CC candidate compound on the basis of selection criteria which include
CC similarities between the atomic coordinates of the selected candidate
CC compound and the atomic coordinates of the mammalian endostatin. The
CC invention also describes the use of an anti-angiogenic fragment of
CC endostatin comprising a domain selected from a heparin binding domain, a
CC receptor binding domain, and exposed on alpha-helix A domain, and a
CC carbohydrate recognition domain (CRD) domain. The methods can be used for
CC designing and selecting endostatin mimics. The compounds identified can
CC be used for treating undesired angiogenesis, e.g. tumours. This sequence
CC represents human alpha1(XVIII) collagen which is used in the description
CC of the method
XX
SQ Sequence 684 AA;

Query Match   79.3%; Score 982; DB 2; Length 684;
Best Local Similarity   81.1%; Pred No. 1.7e-101;
Matches 185; Conservative 15; Mismatches 22; Indels   6; Gaps   3;

Qy   1 PWRADDILAGPRLLDPQPYPGAPHHGSVVH-FQPARPTGGPVHTHTHQDFQLVLHLV 59
         ||||||| ||||| :|||||| :|||||| :|:|:|||
Db 460 PWRADDILASPPGLPEPQPYPG-PHHSSVHTCGPARPTSPP---AHSHRDFQPVLHLV 514

Qy   60 ALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVN 119
         ||||| |||||||||||||||||||| |||||||||||||||||||||| ||:|
Db 515 ALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVN 574

Qy 120 LRDEVLFPSWEALFSGSEGQLKPGARIFSPDGRDVLQHPAWPRKSVWHGSDPSGRRLTDS 179
         |:|||||||||||||||||:||||||||||| ||:|:|| :|::||||||||:|
Db 575 LKDELLFPSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPQKSVWHGSDPNGRRLTES 634

Qy 180 YCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTS 227
         ||||||||||:||||||||||||||||||||::|||||||||||||
Db 635 YCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 682

RESULT 13
AAW26328
ID AAW26328 standard; protein; 1288 AA.
XX
AC AAW26328;
XX
DT 19-NOV-1997 (first entry)
XX

FH Key   Location/Qualifiers
FT Peptide   303..308
FT   /label= GXYGX'Y'_motif
FT Peptide   309..314
FT   /label= GXYGX'Y'_motif
FT Peptide   315..320
FT   /label= GXYGX'Y'_motif
FT Peptide   321..326
FT   /label= GXYGX'Y'_motif
FT Peptide   337..342
FT   /label= GXYGX'Y'_motif
FT Peptide   343..348
FT   /label= GXYGX'Y'_motif
FT Peptide   349..354
FT   /label= GXYGX'Y'_motif
FT Peptide   355..360
FT   /label= GXYGX'Y'_motif
FT Peptide   361..366
FT   /label= GXYGX'Y'_motif
FT Peptide   367..372
FT   /label= GXYGX'Y'_motif
FT Peptide   373..378
FT   /label= GXYGX'Y'_motif
FT Peptide   379..384
FT   /label= GXYGX'Y'_motif
FT Peptide   385..390
FT   /label= GXYGX'Y'_motif
FT Peptide   396..401
FT   /label= GXYGX'Y'_motif
FT Peptide   402..407
FT   /label= GXYGX'Y'_motif
FT Peptide   435..440
FT   /label= GXYGX'Y'_motif
FT Peptide   441..446
FT   /label= GXYGX'Y'_motif
FT Peptide   447..452
FT   /label= GXYGX'Y'_motif
FT Peptide   453..458
FT   /label= GXYGX'Y'_motif
FT Peptide   459..464
FT   /label= GXYGX'Y'_motif
FT Peptide   470..475
FT   /label= GXYGX'Y'_motif
FT Peptide   476..481
FT   /label= GXYGX'Y'_motif
FT Peptide   482..487
FT   /label= GXYGX'Y'_motif
FT Peptide   488..493
FT   /label= GXYGX'Y'_motif
FT Peptide   494..499
FT   /label= GXYGX'Y'_motif
FT Peptide   500..505
FT   /label= GXYGX'Y'_motif
FT Peptide   506..511
FT   /label= GXYGX'Y'_motif
FT Peptide   512..517
FT   /label= GXYGX'Y'_motif
FT Peptide   518..523
FT   /label= GXYGX'Y'_motif
FT Peptide   524..529
FT   /label= GXYGX'Y'_motif
FT Peptide   530..535
FT   /label= GXYGX'Y'_motif
FT Peptide   536..541
FT   /label= GXYGX'Y'_motif
FT Peptide   542..547
FT   /label= GXYGX'Y'_motif

```
FT   Peptide         548..553
FT                   /label= GXYGX'Y'_motif
FT   Peptide         580..585
FT                   /label= GXYGX'Y'_motif
FT   Peptide         586..591
FT                   /label= GXYGX'Y'_motif
FT   Peptide         592..597
FT                   /label= GXYGX'Y'_motif
FT   Peptide         598..603
FT                   /label= GXYGX'Y'_motif
FT   Peptide         604..609
FT                   /label= GXYGX'Y'_motif
FT   Peptide         610..615
FT                   /label= GXYGX'Y'_motif
FT   Peptide         616..621
FT                   /label= GXYGX'Y'_motif
FT   Peptide         622..627
FT                   /label= GXYGX'Y'_motif
FT   Peptide         628..633
FT                   /label= GXYGX'Y'_motif
FT   Peptide         634..639
FT                   /label= GXYGX'Y'_motif
FT   Peptide         640..665
FT                   /label= GXYGX'Y'_motif
FT   Peptide         657..662
FT                   /label= GXYGX'Y'_motif
FT   Peptide         677..682
FT                   /label= GXYGX'Y'_motif
FT   Peptide         683..688
FT                   /label= GXYGX'Y'_motif
FT   Peptide         689..694
FT                   /label= GXYGX'Y'_motif
FT   Peptide         695..700
FT                   /label= GXYGX'Y'_motif
FT   Peptide         707..712
FT                   /label= GXYGX'Y'_motif
FT   Peptide         713..718
FT                   /label= GXYGX'Y'_motif
FT   Peptide         735..740
FT                   /label= GXYGX'Y'_motif
FT   Peptide         741..746
FT                   /label= GXYGX'Y'_motif
FT   Peptide         747..752
FT                   /label= GXYGX'Y'_motif
FT   Peptide         759..764
FT                   /label= GXYGX'Y'_motif
FT   Peptide         765..770
FT                   /label= GXYGX'Y'_motif
FT   Peptide         771..776
FT                   /label= GXYGX'Y'_motif
FT   Peptide         787..792
FT                   /label= GXYGX'Y'_motif
FT   Peptide         793..798
FT                   /label= GXYGX'Y'_motif
FT   Peptide         799..804
FT                   /label= GXYGX'Y'_motif
FT   Peptide         815..820
FT                   /label= GXYGX'Y'_motif
FT   Peptide         821..826
FT                   /label= GXYGX'Y'_motif
FT   Peptide         827..832
FT                   /label= GXYGX'Y'_motif
FT   Peptide         833..838
FT                   /label= GXYGX'Y'_motif
FT   Peptide         839..844
FT                   /label= GXYGX'Y'_motif
FT   Peptide         845..850
FT                   /label= GXYGX'Y'_motif
FT   Peptide         863..868
FT                   /label= GXYGX'Y'_motif
FT   Peptide         869..874
FT                   /label= GXYGX'Y'_motif
FT   Peptide         875..880
FT                   /label= GXYGX'Y'_motif

FT   Peptide         891..896
FT                   /label= GXYGX'Y'_motif
FT   Peptide         897..902
FT                   /label= GXYGX'Y'_motif
FT   Peptide         903..908
FT                   /label= GXYGX'Y'_motif
FT   Peptide         911..916
FT                   /label= GXYGX'Y'_motif
FT   Peptide         917..922
FT                   /label= GXYGX'Y'_motif
FT   Peptide         928..933
FT                   /label= GXYGX'Y'_motif
FT   Peptide         934..939
FT                   /label= GXYGX'Y'_motif
FT   Peptide         956..961
FT                   /label= GXYGX'Y'_motif
FT   Peptide         962..967
FT                   /label= GXYGX'Y'_motif
FT   Peptide         968..973
FT                   /label= GXYGX'Y'_motif
FT   Peptide         1126..1131
FT                   /label= GXYGX'Y'_motif
FT   Peptide         1145..1150
FT                   /label= GXYGX'Y'_motif
FT   Peptide         1193..1198
FT                   /label= GXYGX'Y'_motif
XX
PN   US5643783-A.
XX
PD   01-JUL-1997.
XX
PF   01-DEC-1993;   93US-00159784.
XX
PR   01-DEC-1993;   93US-00159784.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Olsen BR, Oh SP;
XX
DR   WPI; 1997-350247/32.
DR   N-PSDB; AAT84485.
XX
PT   Nucleic acid encoding human alpha-1 collagen - for production of
PT   recombinant alpha-1 collagen, for use in the treatment of cartilage
PT   degeneration.
XX
PS   Disclosure; Fig 2; 35pp; English.

Query Match             79.1%;  Score 980.5;  DB 2;  Length 1288;
Best Local Similarity   79.7%;  Pred. No. 6.3e-101;
Matches  184;  Conservative  19;  Mismatches  23;  Indels  5;  Gaps  2;

Qy      1 PWRADDILAGPRLLDPQYPGAP-HHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLV   59
          ||||||||| || | |||||||  ||||| ||| |||| |||   ||||||||||||
Db   1062 PWRADDILANPPRLPDRQPYPGVPHHHSSYVHLPPARPT---LSLAHTHQDFQPVLHLV 1117

Qy     60 ALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVN  119
          ||:|  ||||||||||||||||||||| |||||||||||||||||||||||| ||:|:||
Db   1118 ALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIVN 1177

Qy    120 LRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDS  179
          | :||| |||  |||| ||:||| |||||||||| | |||||| |||:|||||||| :|
Db   1178 LKDEVLSPSWDSLFSGSGQGVQFGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLMES 1237

Qy    180 YCETWRTEAPAATGQASSLLAGRLLEQRAASCRHAFVVLCIENSVMTSFSK  230
          |||||||| |||||||||||||||||||:|::||| |||||||:::|:|||
Db   1238 YCETWRTETTGATGQASSLLSGRLLEQXAASCHNSYIVLCIENSFMTSFSK 1288


RESULT 14
AAW92297
ID   AAW92297 standard; peptide; 1288 AA.
```

```
XX
AC   AAW92297;
XX
DT   28-APR-1999  (first entry)
XX
DB   Mouse alpha-1 (XVIII) collagen chain common sequence MO18(common)28.
XX
KW   Human; type XVIII collagen; liver disease; cirrhosis; detection;
KW   hepatocellular carcinoma; diagnosis.
OS   Mus sp.
XX
PN   WO9856399-A1.
XX
PD   17-DEC-1998.
XX
PF   12-JUN-1998;   98WO-US012327.
XX
PR   12-JUN-1997;   97US-0049369P.
XX
PA   (FIBR-) FIBROGEN INC.
PA   (FIFI-) ACAD FINLAND.
PA   (INRM ) INST NAT SANTE & RECH MEDICALE.
XX
PI   Pihlajaniemi T,  Rehn M,  Clement B;
XX
DR   WPI; 1999-070292/06.
XX
PT   Diagnosis and monitoring of liver disease by measuring collagen type
PT   XVIII levels - with elevated levels indicative of disease, especially
PT   cirrhosis or hepatocellular carcinoma.
XX
PS   Example 6; Fig 8; 56pp; English.
XX
CC   A method has been developed for the detecting liver disease. The method
CC   comprises: (a) reacting a patient sample with antibodies (Ab) specific
CC   for collagen type XVIII (Col18); (b) measuring the amount of Ab-antigen
CC   complex (C) formed as indicator of the amount of Col18 present; (c)
CC   similar analysis of a non-diseased control; and (d) comparing the amounts
CC   of Col18 in the two samples to detect presence or progression of disease.
CC   Elevated levels of Col18 are: (i) indicative of disease, specifically
CC   cirrhosis; and (ii) predictive of the prognosis of disease, specifically
CC   hepatocellular carcinoma (there is a relationship between Col18 mRNA
CC   levels and tumour size and necrosis, and survival times are significantly
CC   higher in patients with higher Col18 levels). The method provides non-
CC   invasive, early and accurate diagnosis of liver disease. The present
CC   sequence represents the sequence common to mouse alpha-1 (XVIII) collagen
CC   chain from the present invention
XX
SQ   Sequence 1288 AA;

Query Match     78.8%;  Score 976;  DB 2;  Length 1288;
Best Local Similarity   79.7%;  Pred. NO. 2e-100;
Matches  184;  Conservative   20;  Mismatches  21;  Indels   6;  Gaps   3;

Qy      1 PWRADDILAGPPRLLDPQPYPGAP-HHGSYVHF-QPARPTGPVHTHTHQDFQLVLHL  58
            ||||||||||||||||||| :||||  :||||  :           |||||||||||||
Db   1062 PWRADDILANPPRLPDPQPYPGPVPHHHSSYVHLPRPARPT----LSLAHTHQDFQPVLHL 1117

Qy     59 VALNSPQPGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVV  118
            ||||:||:|||||||||||||||:||||||:|||||||||||||||||||:||:
Db   1118 VALNTPLSGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIV 1177

Qy    119 NLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW  166
            |||||||||||||:||||||||||||||||||||||||||||||:
Db   1178 NLKDEVLSPSWDSLFSGSGGQVQPGARIFSFDGRDVLRHPANFQKSVWHGSDPSGGRRLME 1237

Qy    179 SYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFS  229
            ::||||||||||||||:||||:::|||||||   :||||||||||||||||||
Db   1238 SYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMTSFS 1288
```

RESULT 15

```
AAO17430
ID   AAO17430 standard; protein; 184 AA.
XX
AC   AAO17430;
XX
DT   19-JUL-2002  (first entry)
XX
DE   Canine endostatin.
XX
KW   Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;
KW   psoriasis; rheumatoid arthritis; retinopathy; macular degeneration;
KW   corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
KW   rubeosis; Osler-Webber Syndrome; myocardial angiogenesis;
KW   plaque neovascularisation; telangiectasia; haemophiliac joints;
KW   angiofibroma; wound granulation; coronary collateral;
KW   cerebral collateral; arteriovenous malformation;
KW   ischaemic limb angiogenesis; diabetic neovascularisation; fracture;
KW   cytostatic; antiarthritic; antipsoriatic; antidiabetic; ophthalmological;
KW   gynaecological.
XX
OS   Canis familiaris.
XX
PN   EP1191036-A2.
XX
PD   27-MAR-2002.
XX
PF   24-AUG-2001;  2001EP-00307224.
XX
PR   25-AUG-2000;  2000US-0227924P.
XX
PA   (PFIZ ) PFIZER PROD INC.
XX
PI   Sheppard MG,  Tong X;
XX
DR   WPI; 2002-354068/39.
DR   N-PSDB; AAL46063.
XX
PT   An isolated nucleic acid molecule for the treatment of angiogenesis-
PT   related disorder, such as cancers or diabetic retinopathy, encodes an
PT   endostatin protein.
XX
PS   Claim 14; Fig 5; 56pp; English.
XX
CC   The present invention provides the protein and coding sequences of canine
CC   pro-endostatin and endostatin. The sequences can be used in the treatment
CC   and diagnosis of angiogenesis related disorders, including cancer,
CC   rheumatoid arthritis, psoriasis, retinopathy, macular degeneration,
CC   corneal graft rejection, neovascular glaucoma, retrolental fibroplasia,
CC   rubeosis, Osler-Webber Syndrome, myocardial angiogenesis, plaque
CC   neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,
CC   wound granulation, coronary collaterals, cerebral collaterals,
CC   arteriovenous malformations, ischaemic limb angiogenesis, diabetic
CC   neovascularisation, and fractures. The present sequence is the canine pro
CC   -endostatin protein sequence
XX
SQ   Sequence 184 AA;

Query Match     78.0%;  Score 966;  DB 5;  Length 184;
Best Local Similarity  100.0%;  Pred. No. 1.7e-100;
Matches  184;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     47 HTHQDFQLVLHIVALNSPQPGPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI  60

Qy    107 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 166
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120

Qy    167 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
```

us-09-938-391-2.rag

OM protein - protein search, using sw model

Run on: March 26, 2004, 13:34:07 ; Search time 18.3333 Seconds
(without alignments)
647.671 Million cell updates/sec

Title: US-09-938-391-2
Perfect score: 1239
Sequence: 1 PWRADDILAGPPRLLDPQPY..........CRHAFVVLCIENSVWTSFSK 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query* Match	Length	DB	ID	Description
1	843	68.0	191	4	US-09-561-500-13	Sequence 13, Appl
2	843	68.0	191	4	US-09-561-108-13	Sequence 13, Appl
3	843	68.0	191	4	US-09-561-526-13	Sequence 13, Appl
4	843	68.0	191	4	US-09-561-499-13	Sequence 13, Appl
5	843	68.0	191	4	US-09-998-831-13	Sequence 13, Appl
6	839	67.7	195	1	US-08-159-784-2	Sequence 2, Appli
7	835	67.4	182	4	US-09-561-500-14	Sequence 14, Appl
8	835	67.4	182	4	US-09-561-108-14	Sequence 14, Appl
9	835	67.4	182	4	US-09-315-689-3	Sequence 3, Appli
10	835	67.4	182	4	US-09-561-526-14	Sequence 14, Appl
11	835	67.4	182	4	US-09-561-499-14	Sequence 14, Appl
12	835	67.4	182	4	US-09-998-831-14	Sequence 14, Appl
13	835	67.4	183	3	US-09-206-059-2	Sequence 2, Appli
14	817	65.9	178	4	US-09-315-689-5	Sequence 5, Appli
15	790	63.8	185	4	US-08-985-526-36	Sequence 36, Appl
16	513	41.4	191	1	US-08-159-784-3	Sequence 3, Appli
17	195.5	15.8	124	4	US-09-231-077D-10	Sequence 10, Appl
18	188	15.2	123	4	US-09-231-077D-11	Sequence 11, Appl
19	148	11.9	35	3	US-09-046-985-2	Sequence 2, Appli
20	148	11.9	35	3	US-09-474-743-2	Sequence 2, Appli
21	101	8.2	22	3	US-09-046-985-7	Sequence 7, Appli
22	101	8.2	22	3	US-09-474-743-7	Sequence 7, Appli
23	98.5	7.9	369	4	US-09-252-991A-25533	Sequence 25533, A
24	97	7.8	16	3	US-09-185-442-32	Sequence 32, Appl
25	96	7.7	322	4	US-09-252-991A-27326	Sequence 27326, A
26	94.5	7.6	795	4	US-09-252-991A-19085	Sequence 19085, A
27	93	7.5	493	4	US-09-252-991A-23421	Sequence 23421, A
28	91	7.3	618	4	US-09-252-991A-28358	Sequence 28358, A
29	90.5	7.3	757	3	US-09-413-814-84	Sequence 84, Appl
30	90	7.3	534	4	US-09-252-991A-22537	Sequence 22537, A
31	89.5	7.2	432	4	US-09-252-991A-24690	Sequence 24690, A
32	88.5	7.1	630	4	US-09-252-991A-19702	Sequence 19702, A
33	87.5	7.1	541	4	US-09-252-991A-17206	Sequence 17206, A
34	87.5	7.1	816	2	US-08-267-803B-9	Sequence 9, Appli
35	87.5	7.0	816	3	US-09-041-886-17	Sequence 17, Appl
36	87	7.0	478	4	US-09-252-991A-28496	Sequence 28496, A
37	87	7.0	490	4	US-09-252-991A-21383	Sequence 21383, A
38	87	7.0	512	4	US-09-252-991A-17586	Sequence 17586, A
39	87	7.0	731	4	US-09-252-991A-17180	Sequence 17180, A
40	86.5	7.0	420	4	US-09-252-991A-16912	Sequence 16912, A
41	86.5	7.0	654	4	US-09-907-794A-177	Sequence 177, App
42	86.5	7.0	654	4	US-09-905-125A-177	Sequence 177, App
43	86.5	7.0	654	4	US-09-902-775A-177	Sequence 177, App
44	86	6.9	20	2	US-08-740-168A-1	Sequence 1, Appli
45	86	6.9	20	3	US-09-349-429-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-561-500-13

; Sequence 13, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002500
; CURRENT APPLICATION NUMBER: US/09/561,500
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-500-13

Query Match   68.0%; Score 843; DB 4; Length 191;
Best Local Similarity   83.0%; Pred. No. 5.6e-87;
Matches 156; Conservative 17; Mismatches 15; Indels 0; Gaps 0;

```
Qy   43 HTHTHTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQD 102
            | ||||||||||| ||| |||||||||||||||||||| |||||||||||||||||||
Db    4 HHHHHTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLJSGTFRAFLSSRLQD  63

Qy  103 LYSIVRRADRTGVPVVNLRDEVLFPSWEALFSGSEQLKPGARIFSFDGRDVLQHPAWPR  162
            ||||||||| ||| |||||||||||||||||||||||||||||||||:|||||||
Db   64 LYSIVRRADRGSVPIVNLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQ  123

Qy  163 KSVWHGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQBAASCRHAFVVLCIEN  222
            ||||||||||||  ||||:||:||  |||||||||||||||||  :||||||||
Db  124 KSVWHGSDPSGRRLMESYCETWRTETTGATQQASSLLSGRLLEQKAASCHNSYIVLCIEN  183

Qy  223 SVMTSFSK  230
            |||||||
Db  184 SFMTSFSK  191
```

RESULT 2
US-09-561-108-13

; Sequence 13, Application US/09561108
; Patent No. 6342221
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe

```
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/561,108
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
US-09-561-108-13

Query Match             68.0%;  Score 843;  DB 4;  Length 191;
Best Local Similarity   83.0%;  Pred. No. 5.6e-87;
  Matches  156;  Conservative   17;  Mismatches   15;  Indels    0;  Gaps    0;

Qy   43 HTHTHTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQD 102
            |||||||| ||||| ||||  ||||||||||||||||||:|| ||:||||||||||||
Db    4 HHHHHTHQDFQPVLHIVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQD  63

Qy  103 LYSIVRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPR 162
            ||||||||| :|:||:||| :|||| ::|||||:||:|||||||||||||||:||||:
Db   64 LYSIVRRADRGSVPIVNLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQ 123

Qy  163 KSVWHGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIEN 222
            |||||||||||||  ||||||||| || ||||||||:|||||: :::||:: :|||||
Db  124 KSVWHGSDPSGRRLMESYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIEN 183

Qy  223 SVMTSFSK 230
            | |||||
Db  184 SFMTSFSK 191
```

```
RESULT 3
US-09-561-526-13
; Sequence 13, Application US/09561526
; Patent No. 6416758
; GENERAL INFORMATION:
; APPLICANT:  Philip E. Thorpe
; APPLICANT:  Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002586
; CURRENT APPLICATION NUMBER: US/09/561,526
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
US-09-561-526-13

Query Match             68.0%;  Score 843;  DB 4;  Length 191;
Best Local Similarity   83.0%;  Pred. No. 5.6e-87;
  Matches  156;  Conservative   17;  Mismatches   15;  Indels    0;  Gaps    0;

Qy   43 HTHTHTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQD 102
            |||||||| ||||| ||||  ||||||||||||||||||:|| ||:||||||||||||
Db    4 HHHHHTHQDFQPVLHIVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQD  63

Qy  103 LYSIVRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPR 162
            ||||||||| :|:||:||| :|||| ::|||||:||:|||||||||||||||:||||:
Db   64 LYSIVRRADRGSVPIVNLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQ 123

Qy  163 KSVWHGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIEN 222
            |||||||||||||  ||||||||| || ||||||||:|||||: :::||:: :|||||
Db  124 KSVWHGSDPSGRRLMESYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIEN 183

Qy  223 SVMTSFSK 230
            | |||||
Db  184 SFMTSFSK 191
```

```
RESULT 4
US-09-561-499-13
; Sequence 13, Application US/0956l499
; Patent No. 652583
; GENERAL INFORMATION:
; APPLICANT:  Philip E. Thorpe
; APPLICANT:  Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/09/561,499
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
US-09-561-499-13

Query Match             68.0%;  Score 843;  DB 4;  Length 191;
Best Local Similarity   83.0%;  Pred. No. 5.6e-87;
  Matches  156;  Conservative   17;  Mismatches   15;  Indels    0;  Gaps    0;

Qy   43 HTHTHTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQD 102
            |||||||| ||||| ||||  ||||||||||||||||||:|| ||:||||||||||||
Db    4 HHHHHTHQDFQPVLHIVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQD  63

Qy  103 LYSIVRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPR 162
            ||||||||| :|:||:||| :|||| ::|||||:||:|||||||||||||||:||||:
Db   64 LYSIVRRADRGSVPIVNLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQ 123

Qy  163 KSVWHGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIEN 222
            |||||||||||||  ||||||||| || ||||||||:|||||: :::||:: :|||||
Db  124 KSVWHGSDPSGRRLMESYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIEN 183

Qy  223 SVMTSFSK 230
            | |||||
Db  184 SFMTSFSK 191
```

```
RESULT 5
US-09-998-831-13
; Sequence 13, Application US/09998831
; Patent No. 6675941
; GENERAL INFORMATION:
; APPLICANT:  Philip E. Thorpe
; APPLICANT:  Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/998,831
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 09/561,108
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-998-831-13

Query Match   68.0%;   Score 843;   DB 4;   Length 191;
Best Local Similarity   83.0%;   Pred. No. 5.6e-87;
Matches 156; Conservative   17;   Mismatches   15;   Indels   0;   Gaps   0;

Qy   43 HTHTHTHQDFQJLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQD 102
         | ||||||||||||:||| |||||||||||||||||||||||||||||||||||||||
Db   4 HHHHHTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQD 63

Qy 103 LYSIVRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPR 162
         ||||||||| |:|||:|||||||||:|||||:||||||||||||||||||:||||||:
Db   64 LYSIVRRADRGSVPIVNLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQ 123

Qy 163 KSVWHGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIEN 222
         ||||||||||||||:||||||||||||| ||||||||||||||||| :::|||||||
Db 124 KSVWHGSDPSGRRLMESYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIEN 183

Qy 223 SVMTSFSK 230
         |||||||
Db 184 SFMTSFSK 191

RESULT 6
US-08-159-784-2
;   Sequence 2, Application US/08159784
;   Patent No. 5643783
;   GENERAL INFORMATION:
;   APPLICANT: Bjorn R. Olsen
;   TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
;   NUMBER OF SEQUENCES: 9
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: Massachusetts
;   COUNTRY: U.S.A.
;   ZIP: 02110-2804
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;   COMPUTER: IBM PS/2 Model 50Z or 55SX
;   OPERATING SYSTEM: MS-DOS (Version 5.0)
;   SOFTWARE: WordPerfect (Version 5.1)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/159,784
;   FILING DATE: December 1, 1993
;   CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: John F. Freeman
;   REGISTRATION NUMBER: 29,066
;   REFERENCE/DOCKET NUMBER: 00246/170001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 542-5070
;   TELEFAX: (617) 542-8906
;   TELEX: 200154
;   INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 195
;   TYPE: amino acid
;   STRANDEDNESS: N/A
;   TOPOLOGY: N/A
US-08-159-784-2

Query Match   67.7%;   Score 839;   DB 1;   Length 195;
Best Local Similarity   80.2%;   Pred. No. 1.6e-86;
Matches 158; Conservative   19;   Mismatches   16;   Indels   4;   Gaps   1;

Qy   34 PARPTGGPVHTHTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFR 93

Db   3 PARPT----LSLAHTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFR 58

Qy   94 AFLSSRLQDLYSIVRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRD 153
         ||||||||||||||||||| |:|||:||||||||||:|||||:||||||:||||||||||
Db   59 AFLSSRLQDLYSIVRRADRGSVPIVNLKDEVLSPSWDSLFSGSQGQVQPGARIFSFDGRD 118

Qy 154 VLCHPAWPRKSVWHGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRH 213
         ||||||| ||||||||||||||:|||||||||||| ||||||||||||||| :::||
Db 119 VLRHPAWPQKSVWHGSDPSGRRLMESYCETWRTETTGATGQASSLLSGRLLEQKAASCHN 178

Qy 214 AFVVLCIENSVMTSFSK 230
         :::||||||||||||||
Db 179 SYIVLCIENSFMTSFSK 195

RESULT 7
US-09-561-500-14
;   Sequence 14, Application US/09561500
;   Patent No. 6342219
;   GENERAL INFORMATION:
;   APPLICANT: Philip E. Thorpe
;   APPLICANT: Rolf A. Brekken
;   TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
;   FILE REFERENCE: 4001.002500
;   CURRENT APPLICATION NUMBER: US/09/561,500
;   CURRENT FILING DATE: 2000-04-28
;   PRIOR APPLICATION NUMBER: 60/131,432
;   PRIOR FILING DATE: 1999-04-28
;   NUMBER OF SEQ ID NOS: 44
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 14
;   LENGTH: 182
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-561-500-14

Query Match   67.4%;   Score 835;   DB 4;   Length 182;
Best Local Similarity   85.1%;   Pred. No. 4.1e-86;
Matches 154; Conservative   14;   Mismatches   13;   Indels   0;   Gaps   0;

Qy   47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
         |||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db   1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy 107 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 166
         ||||||: |:|||:|||||||||:|||||:||||||||||||||||:||||||:|:|||
Db   61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy 167 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
         ||||||||||:|||||||||||| ||||||||||||||||| ::::|||||||||||
Db 121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy 227 S 227
         :
Db 181 A 181

RESULT 8
US-09-561-108-14
;   Sequence 14, Application US/09561108
;   Patent No. 6342221
;   GENERAL INFORMATION:
;   APPLICANT: Philip E. Thorpe
;   APPLICANT: Rolf A. Brekken
;   TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
;   FILE REFERENCE: 4001.002584
;   CURRENT APPLICATION NUMBER: US/09/561,108
;   CURRENT FILING DATE: 2000-04-28
;   PRIOR APPLICATION NUMBER: 60/131,432

; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-561-108-14

```
Query Match             67.4%;  Score 835;  DB 4;  Length 182;
Best Local Similarity   85.1%;  Pred. No. 4.1e-86;
Matches   154; Conservative   14; Mismatches   13; Indels    0; Gaps    0;

Qy   47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
           |:::||:|||||||||||  ||  |||||||||||||| |||||||||||||||||||||
Db    1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI  60

Qy  107 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 166
           ||||   ||:||||:|:|||||||||||||||:|||||||||||:|||::||
Db   61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy  167 HGSDPSGRRLTDSYCETWRTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
           |||||:||||||:|||||||||::||||||||:||||: |:|||:|||:|||||||:||
Db  121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy  227 S 227
          :
Db  181 A 181
```

RESULT 9
US-09-315-689-3
; Sequence 3, Application US/09315689
; Patent No. 6346510
; GENERAL INFORMATION:
; APPLICANT: Folkman, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
; FILE REFERENCE: 05213-0229
; CURRENT APPLICATION NUMBER: US/09/315,689
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-315-689-3

```
Query Match             67.4%;  Score 835;  DB 4;  Length 182;
Best Local Similarity   85.1%;  Pred. No. 4.1e-86;
Matches   154; Conservative   14; Mismatches   13; Indels    0; Gaps    0;

Qy   47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
           |:::||:|||||||||||  ||  |||||||||||||| |||||||||||||||||||||
Db    1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI  60

Qy  107 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 166
           ||||   ||:||||:|:|||||||||||||||:|||||||||||:|||::||
Db   61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy  167 HGSDPSGRRLTDSYCETWRTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
           |||||:||||||:|||||||||::||||||||:||||: |:|||:|||:|||||||:||
Db  121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy  227 S 227
          :
Db  181 A 181
```

RESULT 10
US-09-561-526-14
; Sequence 14, Application US/09561526
; Patent No. 6416758
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002586
; CURRENT APPLICATION NUMBER: US/09/561,526
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-561-526-14

```
Query Match             67.4%;  Score 835;  DB 4;  Length 182;
Best Local Similarity   85.1%;  Pred. No. 4.1e-86;
Matches   154; Conservative   14; Mismatches   13; Indels    0; Gaps    0;

Qy   47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
           |:::||:|||||||||||  ||  |||||||||||||| |||||||||||||||||||||
Db    1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI  60

Qy  107 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 166
           ||||   ||:||||:|:|||||||||||||||:|||||||||||:|||::||
Db   61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy  167 HGSDPSGRRLTDSYCETWRTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
           |||||:||||||:|||||||||::||||||||:||||: |:|||:|||:|||||||:||
Db  121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy  227 S 227
          :
Db  181 A 181
```

RESULT 11
US-09-561-499-14
; Sequence 14, Application US/09561499
; Patent No. 6524583
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/09/561,499
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-561-499-14

```
Query Match             67.4%;  Score 835;  DB 4;  Length 182;
Best Local Similarity   85.1%;  Pred. No. 4.1e-86;
Matches   154; Conservative   14; Mismatches   13; Indels    0; Gaps    0;
```

```
Qy    47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
         |:|:||| |||||||||||  |||:||||||||||||||||||||||||||||||||||
Db     1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy   107 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 166
         |||||| |:|||||||||||  ||:||||||| |||||||||||||| |:|:||||
Db    61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy   167 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
         ||||| ||||||||||||||||| |:|||||||||||||||||  || ||:||||||||
Db   121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy   227 S 227
         :
Db   181 A 181

RESULT 12
US-09-998-831-14
; Sequence 14, Application US/09998831
; Patent No. 6676941
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
; TITLE OF INVENTION: INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/998,831
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 09/561,108
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-998-831-14

Query Match             67.4%;    Score 835;  DB 4;  Length 182;
Best Local Similarity   85.1%;    Pred. No. 4.1e-86;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Qy    47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
         |:|:||| |||||||||||  |||:||||||||||||||||||||||||||||||||||
Db     1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy   107 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 166
         |||||| |:|||||||||||  ||:||||||| |||||||||||||| |:|:||||
Db    61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy   167 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
         ||||| ||||||||||||||||| |:|||||||||||||||||  || ||:||||||||
Db   121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy   227 S 227
         :
Db   181 A 181

RESULT 13
US-09-206-059-2
; Sequence 2, Application US/09206059
; Patent No. 6201104
; GENERAL INFORMATION:
; APPLICANT: MacDonald, Nicholas
; APPLICANT: Sim, Kim Lee
; TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and
; TITLE OF INVENTION: Proteins and Methods of Use
; FILE REFERENCE: 05213-0370
; CURRENT APPLICATION NUMBER: US/09/206,059
; CURRENT FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-206-059-2

Query Match             67.4%;    Score 835;  DB 3;  Length 183;
Best Local Similarity   85.1%;    Pred. No. 4.2e-86;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Qy    47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
         |:|:||| |||||||||||  |||:||||||||||||||||||||||||||||||||||
Db     1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy   107 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 166
         |||||| |:|||||||||||  ||:||||||| |||||||||||||| |:|:||||
Db    61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy   167 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
         ||||| ||||||||||||||||| |:|||||||||||||||||  || ||:||||||||
Db   121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy   227 S 227
         :
Db   181 A 181

RESULT 14
US-09-315-689-5
; Sequence 5, Application US/09315689
; Patent No. 6346510
; GENERAL INFORMATION:
; APPLICANT: Folkman, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
; FILE REFERENCE: 05213-0229
; CURRENT APPLICATION NUMBER: US/09/315,689
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-315-689-5

Query Match             65.9%;    Score 817;  DB 4;  Length 178;
Best Local Similarity   85.9%;    Pred. No. 4.3e-84;
Matches 152; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

Qy    51 DFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRA 110
         ||| |:|:|||||||||  ||:|||||||||||||||||||||||||||||||| |||
Db     1 DFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRA 60

Qy   111 DRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSD 170
         || |:|||||||||||  ||:||||||| |||||||||||||| |:|:|||| ||||
Db    61 DRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSD 120

Qy   171 PSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTS 227
         | ||||||||||||||| |:|||||||||||||||:  || ||:||||||||||
Db   121 PNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 177

RESULT 15
US-08-985-526-36
; Sequence 36, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
;
```

TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
TITLE OF INVENTION: THERAPY
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
    ADDRESSEE: Connolly, Bove, Lodge, & Hutz
    STREET: 1220 Market Street, P.O. Box 2207
    CITY: Wilmington
    STATE: Delaware
    COUNTRY: U.S.A.
    ZIP: 19899
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/985,526
    FILING DATE:
    CLASSIFICATION:
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/608,845
    FILING DATE: 16-JUL-1996
ATTORNEY/AGENT INFORMATION:
    NAME: McMorrow Jr., Robert G
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (302) 658-9141
    TELEFAX: (302) 658-5613
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
    LENGTH: 185 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
US-08-985-526-36

Query Match   63.8%;   Score 790; DB 3; Length 185;
Best Local Similarity 81.1%;   Pred. No. 5e-81;
Matches 150; Conservative 16; Mismatches 17; Indels   2; Gaps   2;

Qy   47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
          ||||||||||||||||| |||||||||||||||||| ||:|||||||||||||||||||
Db   2 HTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCFNNAR-VGLSGTFRAFLSSRLQDLYSI 60

Qy   107 VRRADRTGVPVV-NLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSV 165
          ||||||:|:| : |||||||| |||||||:||::|||||||||||||:|||||||:|:|
Db   61 VRRADRGSVPIVQNLRDEVLSPSWDSLPSGSQGQLQPGARIFSFDGRDVLRHPAWPQRSV 120

Qy   166 WHGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 225
          ||||||||||||| :||||||||| ||||||||:|||||| :::||||||| : ||||||
Db   121 WHGSDPSGRRLMESYCETWRTETTGATGQASSLLSGRLLEQRAASCHDSYIVLCIENSFM 180

Qy   226 TSFSK 230
          ||||:
Db   181 TSFSR 185

Search completed: March 26, 2004, 13:39:22
Job time : 20.3333 secs

```
                  Copyright   GenCore version 5.1.6
                  Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:           March 26, 2004, 13:38:08 ; Search time 41.6667 Seconds
                                             (without alignments)
                                             1444.373 Million cell updates/sec

Title:            US-09-938-391-2
Perfect score:    1239
Sequence:         1 PWRADDILAGPPRLLDPQPY.........CRHAFVVLCIENSVMTSFSK 230

Scoring table:    BLOSUM62
                  Gapop 10.0 , Gapext 0.5

Searched:         1065169 seqs, 261661801 residues

Total number of hits satisfying chosen parameters:        1065169

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published Applications AA:*
   1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
   2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
   3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
   4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
   5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
   6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
   7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
   8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
   9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
  10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
  11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
  12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
  13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
  14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
  15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
  16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
  17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
  18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	1239	100.0	230	10	US-09-938-391-2	Sequence 2, Appli
2	1019	82.2	682	15	US-10-264-049-3010	Sequence 3010, Ap
3	1019	82.2	1516	14	US-10-060-036-166	Sequence 166, App
4	1019	82.2	1516	15	US-10-431-642-3	Sequence 3, Appli
5	1003	81.0	684	10	US-09-961-403-5	Sequence 5, Appli
6	966	78.0	184	10	US-09-938-391-4	Sequence 4, Appli
7	959	77.4	184	14	US-10-131-241-49	Sequence 49, Appl
8	959	77.4	184	14	US-10-292-418-35	Sequence 35, Appl
9	843	68.0	191	9	US-09-998-831-13	Sequence 13, Appl
10	843	68.0	191	14	US-10-373-561-13	Sequence 13, Appl
11	835	67.4	181	14	US-10-131-241-55	Sequence 55, Appl
12	835	67.4	182	9	US-09-998-831-14	Sequence 14, Appl
13	835	67.4	182	14	US-10-131-241-54	Sequence 54, Appl
14	835	67.4	182	14	US-10-042-347-3	Sequence 3, Appli
15	835	67.4	182	14	US-10-373-561-14	Sequence 14, Appl
16	835	67.4	183	9	US-09-873-676-2	Sequence 2, Appli
17	835	67.4	183	13	US-10-080-797-1	Sequence 1, Appli
18	835	67.4	183	14	US-10-131-241-52	Sequence 52, Appl
19	835	67.4	183	14	US-10-292-418-4	Sequence 4, Appl
20	834	67.3	180	14	US-10-131-241-56	Sequence 56, Appl
21	831	67.1	180	14	US-10-131-241-47	Sequence 47, Appl
22	831	67.1	207	13	US-10-080-797-3	Sequence 3, Appli
23	828	66.8	184	14	US-10-131-241-46	Sequence 46, Appl
24	828	66.8	184	14	US-10-292-418-18	Sequence 18, Appl
25	828	66.8	207	14	US-10-422-934-71	Sequence 71, Appl
26	822	66.3	184	12	US-10-210-172-162	Sequence 162, App
27	817	65.9	178	12	US-10-131-241-60	Sequence 60, Appl
28	817	65.9	178	14	US-10-042-347-5	Sequence 5, Appli
29	817	65.9	179	14	US-10-131-241-57	Sequence 57, Appl
30	790	63.8	185	13	US-10-036-869-36	Sequence 36, App
31	597.5	48.2	160	12	US-10-210-172-164	Sequence 164, App
32	504	40.7	160	12	US-10-210-172-166	Sequence 166, App
33	302	24.4	63	9	US-09-822-540A-1	Sequence 1, Appli
34	161	13.0	31	9	US-09-822-540A-2	Sequence 2, Appli
35	103	8.3	24	15	US-10-016-569A-29	Sequence 29, Appl
36	103	8.3	24	15	US-10-308-644-29	Sequence 29, Appl
37	100.5	8.1	739	16	US-10-156-761-13097	Sequence 13097, A
38	97	7.8	16	9	US-09-766-412-32	Sequence 32, Appl
39	91	7.3	431	15	US-10-108-260A-3749	Sequence 3749, Ap
40	91	7.3	6239	15	US-10-156-761-8477	Sequence 8477, Ap
41	89.5	7.2	437	12	US-10-425-114-42857	Sequence 42857, A
42	89	7.2	517	12	US-10-425-114-41375	Sequence 41375, A
43	88	7.1	3436	14	US-10-132-134-32	Sequence 32, Appl
44	87.5	7.1	816	14	US-10-207-706-3	Sequence 3, Appli
45	87	7.0	19725	15	US-10-084-846A-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT 1
US-09-938-391-2
; Sequence 2, Application US/09938391
; Publication No. US20030158099A1
; GENERAL INFORMATION:
;   APPLICANT: Tong, et al.
;   TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND TREATING
;   TITLE OF INVENTION: DISORDERS INVOLVING ANGIOGENESIS
;   FILE REFERENCE: PC10790A
;   CURRENT APPLICATION NUMBER: US/09/938,391
;   CURRENT FILING DATE: 2001-08-24
;   NUMBER OF SEQ ID NOS: 4
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 230
;   TYPE: PRT
;   ORGANISM: CANINE PRO ENDOSTATIN AMINO ACID SEQUENCE
US-09-938-391-2

  Query Match             100.0%;  Score 1239;  DB 10;  Length 230;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-118;
  Matches  230;  Mismatches   0;  Conservative   0;  Indels   0;  Gaps   0;

Qy        1 PWRADDILAGPPRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLVA 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 PWRADDILAGPPRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLVA 60

Qy       61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120

Qy      121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180

Qy      181 CETWRTEAPAATCQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 230
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 CETWRTEAPAATCQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 230
```

RESULT 2
US-10-264-049-3010
; Sequence 3010, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
;   APPLICANT: Birse et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PA133P1
;   CURRENT APPLICATION NUMBER: US/10/264,049
;   CURRENT FILING DATE: 2002-10-04
;   PRIOR APPLICATION NUMBER: PCT/US01/18569
;   PRIOR FILING DATE: 2001-06-07
;   PRIOR APPLICATION NUMBER: US 60/209,467
;   PRIOR FILING DATE: 2000-06-07
;   NUMBER OF SEQ ID NOS: 4360
;   SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 3010
;   LENGTH: 682
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (20)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (39)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-3010

Query Match   82.2%;   Score 1019;   DB 15;   Length 682;
Best Local Similarity   82.4%;   Pred. No. 2.1e-95;
Matches   187;   Conservative   16;   Mismatches   20;   Indels   4;   Gaps   1;

Qy   1 PWRADDILAGPPRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLVA   60
             ||||||||:||||||||||||||||||||:||.:||| ||:||:|||||||
Db   458 PWRADDILASPPRLPEPQPYPGAPHHSSYVHLRPARPTSPP---AHSHRDFQPVLHLVA   513

Qy   61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL   120
             ||||.:||||||||||||||||||||||||||||||||||||||||||||||||:.:||
Db   514 LNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL   573

Qy   121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY   180
             :||:|||||||||||||||||||||||||||||.|:||:||||.|||||||.:|||::|
Db   574 KDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTESY   633

Qy   181 CETWRTEAPAATGQASLLAGRLLEQBAASCRHAFVVLCIENSVMTS   227
             |||||||||:||||||||||||||||:||||||||||||||:|||
Db   634 CETWRTEAPSATGQASLLGGRLLGQSAASCHHAYIVLCIENSFMTA   680

RESULT 3
US-10-060-036-166
; Sequence 166, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
;   APPLICANT: Benson, Darin R.
;   APPLICANT: Kalos, Michael D.
;   APPLICANT: Lodes, Michael J.
;   APPLICANT: Persing, David H.
;   APPLICANT: Hepler, William T.
;   APPLICANT: Jiang, Yuqiu
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;   TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
;   FILE REFERENCE: 210121.566
;   CURRENT APPLICATION NUMBER: US/10/060,036
;   CURRENT FILING DATE: 2002-01-30
;   NUMBER OF SEQ ID NOS: 4560
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 166
;   LENGTH: 1516

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-060-036-166

Query Match   82.2%;   Score 1019;   DB 14;   Length 1516;
Best Local Similarity   82.4%;   Pred. No. 5.9e-95;
Matches   187;   Conservative   16;   Mismatches   20;   Indels   4;   Gaps   1;

Qy   1 PWRADDILAGPPRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLVA   60
             ||||||||:||||||||||||||||||||:||.:||| ||:||:|||||||
Db   1292 PWRADDILASPPRLPEPQPYPGAPHHSSYVHLRPARPTSPP---AHSHRDFQPVLHLVA 1347

Qy   61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL   120
             ||||.:||||||||||||||||||||||||||||||||||||||||||||||||:.:||
Db   1348 LNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL 1407

Qy   121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY   180
             :||:|||||||||||||||||||||||||||||.|:||:||||.|||||||.:|||::|
Db   1408 KDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTESY 1467

Qy   181 CETWRTEAPAATGQASLLAGRLLEQBAASCRHAFVVLCIENSVMTS   227
             |||||||||:||||||||||||||||:||||||||||||||:|||
Db   1468 CETWRTEAPSATGQASLLGGRLLGQSAASCHHAYIVLCIENSFMTA 1514

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-060-036-166

Query Match   82.2%;   Score 1019;   DB 14;   Length 1516;
Best Local Similarity   82.4%;   Pred. No. 5.9e-95;
Matches   187;   Conservative   16;   Mismatches   20;   Indels   4;   Gaps   1;

Qy   1 PWRADDILAGPPRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLVA   60
             ||||||||:||||||||||||||||||||:||.:||| ||:||:|||||||
Db   1292 PWRADDILASPPRLPEPQPYPGAPHHSSYVHLRPARPTSPP---AHSHRDFQPVLHLVA 1347

Qy   61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL   120
             ||||.:||||||||||||||||||||||||||||||||||||||||||||||||:.:||
Db   1348 LNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL 1407

Qy   121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY   180
             :||:|||||||||||||||||||||||||||||.|:||:||||.|||||||.:|||::|
Db   1408 KDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTESY 1467

Qy   181 CETWRTEAPAATGQASLLAGRLLEQBAASCRHAFVVLCIENSVMTS   227
             |||||||||:||||||||||||||||:||||||||||||||:|||
Db   1468 CETWRTEAPSATGQASLLGGRLLGQSAASCHHAYIVLCIENSFMTA 1514

RESULT 4
US-10-431-642-3
; Sequence 3, Application US/10431642
; Publication No. US20040009920A1
; GENERAL INFORMATION:
;   APPLICANT: Ruoslahti, Erkki
;   TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING
;   TITLE OF INVENTION: TUMOR GROWTH AND ANGIOGENESIS
;   FILE REFERENCE: BURNHAM.008CIP
;   CURRENT APPLICATION NUMBER: US/10/431,642
;   CURRENT FILING DATE: 2003-05-05
;   PRIOR APPLICATION NUMBER: 10/005,171
;   PRIOR FILING DATE: 2001-12-03
;   PRIOR APPLICATION NUMBER: 60/331,357
;   PRIOR FILING DATE: 2000-12-04
;   NUMBER OF SEQ ID NOS: 4
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 1516
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-10-431-642-3

RESULT 5
US-09-961-403-5
; Sequence 5, Application US/09961403
; Publication No. US20030077589A1

; GENERAL INFORMATION:
; APPLICANT: HE-STUMPP, HOLGER
; APPLICANT: HAENDLER, BERNARD
; APPLICANT: KRAETZSCHMAR, JOERN
; APPLICANT: KREFT, BERTHOLT
; APPLICANT: WINTERHAGER, ELKE
; APPLICANT: REGIDOR, PEDRO
; APPLICANT: SCOTTI, SIMONE
; TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS
; FILE REFERENCE: SCH-1789
; CURRENT APPLICATION NUMBER: US/09/961,403
; CURRENT FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 5
; LENGTH: 684
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-403-5

Query Match   81.0%;   Score 1003; DB 10; Length 684;
Best Local Similarity   81.5%;   Pred. No. 9.1e-94;
Matches 185; Conservative 15; Mismatches 23; Indels 4; Gaps 1;

Qy   1 PWRADDILAGPPRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHQDPQLVLHLVA   60
          |||||||| |:|||||| ||| || : ||:||| ||||||||
Db   460 PWRADDILASPPGLPEPQPYPGGPHHSSYVHCGPARPTSPP----AHSHRDFQPVLHLVA 515

Qy   61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
          ||||| ||||||||||||||||||||:|||||||||||||||||||| |:|||
Db   516 LNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL 575

Qy   121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180
          ||:|||||||||||||||||||||||||||||||| ||:||:||||||||||
Db   576 KDELFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTESY 635

Qy   181 CETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTS 227
          ||||||||:|||||||:|| ||||| :: ||:||||||||
Db   636 CETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLVLCIENSFMTA 682

RESULT 6
US-09-938-391-4
; Sequence 4, Application US/09938391
; Publication No. US20030158099A1
; GENERAL INFORMATION:
; APPLICANT: Tong, et al.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND TREATING
; TITLE OF INVENTION: DISORDERS INVOLVING ANGIOGENESIS
; FILE REFERENCE: PC10790A
; CURRENT APPLICATION NUMBER: US/09/938,391
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 4
; LENGTH: 184
; TYPE: PRT
; ORGANISM: CANINE ENDOSTATIN AMINO ACID SEQUENCE
US-09-938-391-4

Query Match   78.0%;   Score 966; DB 10; Length 184;
Best Local Similarity 100.0%;   Pred. No. 1e-90;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI   60

Qy   107 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 166
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120

Qy   167 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180

Qy   227 SFSK 230
          ||||
Db   181 SFSK 184

RESULT 7
US-10-131-241-49
; Sequence 49, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolife
; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
;
; SEQ ID NO 49
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Canine sp.
US-10-131-241-49

Query Match   77.4%;   Score 959; DB 14; Length 184;
Best Local Similarity   99.5%;   Pred. No. 5.3e-90;
Matches 183; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 HTHQDFQPVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI   60

Qy   107 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 166
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120

Qy   167 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180

Qy   227 SFSK 230
          ||||
Db   181 SFSK 184

RESULT 8
US-10-292-418-35
; Sequence 35, Application US/10292418
; Publication No. US20030139365A1
; GENERAL INFORMATION:
; APPLICANT: Lo, Kin-Ming
; APPLICANT: Li, Yue
; APPLICANT: Gillies, Stephen D
; TITLE OF INVENTION: Expression and Export of Angiogenesis Inhibitors as
; TITLE OF INVENTION: Immunofusins
; FILE REFERENCE: LEX-006C1
; CURRENT APPLICATION NUMBER: US/10/292,418
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: 09/383,315
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: US 60/097,883
; PRIOR FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 54

```
Query Match             77.4%;   Score 959;  DB 14;   Length 184;
Best Local Similarity   99.5%;   Pred. No. 5.3e-90;
Matches   183;   Conservative   0;   Mismatches   1;   Indels   0;   Gaps   0;

Qy    47  HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  HTHQDFQPVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI  60

Qy   107  VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120

Qy   167  HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180

Qy   227  SFSK 230
              ||||
Db   181  SFSK 184
```

RESULT 9

```
Query Match             68.0%;   Score 843;  DB 9;   Length 191;
Best Local Similarity   83.0%;   Pred. No. 4.3e-78;
Matches   156;   Conservative   17;   Mismatches   15;   Indels   0;   Gaps   0;

Qy    43  HTHTHTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQD 102
              ||||||| |||||||||||  |:|||||||||||||||||| |:|:|||||||||||||
Db     4  HHHHHTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQD  63

Qy   103  LYSIVRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPR 162
              ||||||||||  |:||||:||||  ||: :|||||: ||:|||||||||||||:||||:
Db    64  LYSIVRRADRGSVPIVNLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQ 123

Qy   163  KSVWHGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIEN 222
              ||||||||||||| :|||||||||  ::|||||||:|||||||:|||||:::|:|||||
Db   124  KSVWHGSDPSGRRLMESYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIEN 183

Qy   223  SVMTSFSK 230
              || |||||
Db   184  SFMTSFSK 191
```

RESULT 10

```
Query Match             68.0%;   Score 843;  DB 14;   Length 191;
Best Local Similarity   83.0%;   Pred. No. 4.3e-78;
Matches   156;   Conservative   17;   Mismatches   15;   Indels   0;   Gaps   0;

Qy    43  HTHTHTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQD 102
              ||||||| |||||||||||  |:|||||||||||||||||| |:|:|||||||||||||
Db     4  HHHHHTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQD  63

Qy   103  LYSIVRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPR 162
              ||||||||||  |:||||:||||  ||: :|||||: ||:|||||||||||||:||||:
Db    64  LYSIVRRADRGSVPIVNLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQ 123

Qy   163  KSVWHGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIEN 222
              ||||||||||||| :|||||||||  ::|||||||:|||||||:|||||:::|:|||||
Db   124  KSVWHGSDPSGRRLMESYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIEN 183

Qy   223  SVMTSFSK 230
              || |||||
Db   184  SFMTSFSK 191
```

RESULT 11

```
Query Match             67.4%;   Score 835;  DB 14;   Length 181;
```

TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolife
TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
FILE REFERENCE: 05213-0344   43170-271565
CURRENT APPLICATION NUMBER: US/10/131,241
CURRENT FILING DATE: 2002-07-22
PRIOR APPLICATION NUMBER: US 09/413,049
PRIOR FILING DATE: 1999-10-06
PRIOR APPLICATION NUMBER: US 09/316,802
PRIOR FILING DATE: 1999-05-21
PRIOR APPLICATION NUMBER: US 60/086,586
PRIOR FILING DATE: 1998-05-22
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn version 3.1
SEQ ID NO 54
LENGTH: 182
TYPE: PRT
ORGANISM: Homo sapiens
US-10-131-241-54

```
Query Match              67.4%;  Score 835;  DB 14;  Length 182;
Best Local Similarity 85.1%;  Pred. No. 2.7e-77;
Matches 154;  Conservative  14;  Mismatches  13;  Indels  0;  Gaps  0;

Qy   47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
           :|||||||||||||||:||||||||||||||||||:||||||||||||||
Db    1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI  60

Qy  107 VRRADRTGVPVVNLRDEVLFPSWEALPSGSEGQLKPGARIFSPDGRDVLQHPAWPRKSVW 166
           ||||:||:||:||:||||||||||||||||:||||||||||||||:||:||:|:|:||
Db   61 VRRADRAAVPIVNLKDELLFPSWEALPSGSEGPLKPGARIFSPDGKDVLRHPTWPQKSVW 120

Qy  167 HGSDPSGRRLTDSYCETWRTEAPAATGQASLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
           ||||:|||:|||:||||||||||:|||||||:||:|||||  :::|||||||||| ||
Db  121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy  227 S 227
          :
Db  181 A 181
```

RESULT 14
US-10-042-347-3
; Sequence 3, Application US/10042347
; Publication No. US20030114370A1
; GENERAL INFORMATION:
; APPLICANT: O'Reilly, Michael S.
; APPLICANT: Folkman, M. Judah
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Endostatin Protein and Peptide F
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 05213-0880   (43170-249874)
; CURRENT APPLICATION NUMBER: US/10/042,347
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: US 09/315,689
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: US 60/106,343
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: US 09/154,302
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: US 08/740,168
; PRIOR FILING DATE: 1996-10-22
; PRIOR APPLICATION NUMBER: US 60/005,835
; PRIOR FILING DATE: 1995-10-23
; PRIOR APPLICATION NUMBER: US 60/023,070
; PRIOR FILING DATE: 1996-08-02
; PRIOR APPLICATION NUMBER: US 50/026,263
; PRIOR FILING DATE: 1996-09-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-347-3

```
Best Local Similarity 85.1%;  Pred. No. 2.7e-77;
Matches 154;  Conservative  14;  Mismatches  13;  Indels  0;  Gaps  0;

Qy   47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
           :|||||||||||||||:||||||||||||||||||:||||||||||||||
Db    1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI  60

Qy  107 VRRADRTGVPVVNLRDEVLFPSWEALPSGSEGQLKPGARIFSPDGRDVLQHPAWPRKSVW 166
           ||||:||:||:||:||||||||||||||||:||||||||||||||:||:||:|:|:||
Db   61 VRRADRAAVPIVNLKDELLFPSWEALPSGSEGPLKPGARIFSPDGKDVLRHPTWPQKSVW 120

Qy  167 HGSDPSGRRLTDSYCETWRTEAPAATGQASLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
           ||||:|||:|||:||||||||||:|||||||:||:|||||  :::|||||||||| ||
Db  121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy  227 S 227
          :
Db  181 A 181
```

RESULT 12
US-09-998-831-14
; Sequence 14, Application US/09998831
; Patent No. US20020191953A1
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
; TITLE OF INVENTION: INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/998,831
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 09/561,108
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-998-831-14

```
Query Match              67.4%;  Score 835;  DB 9;  Length 182;
Best Local Similarity 85.1%;  Pred. No. 2.7e-77;
Matches 154;  Conservative  14;  Mismatches  13;  Indels  0;  Gaps  0;

Qy   47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
           :|||||||||||||||:||||||||||||||||||:||||||||||||||
Db    1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI  60

Qy  107 VRRADRTGVPVVNLRDEVLFPSWEALPSGSEGQLKPGARIFSPDGRDVLQHPAWPRKSVW 166
           ||||:||:||:||:||||||||||||||||:||||||||||||||:||:||:|:|:||
Db   61 VRRADRAAVPIVNLKDELLFPSWEALPSGSEGPLKPGARIFSPDGKDVLRHPTWPQKSVW 120

Qy  167 HGSDPSGRRLTDSYCETWRTEAPAATGQASLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
           ||||:|||:|||:||||||||||:|||||||:||:|||||  :::|||||||||| ||
Db  121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy  227 S 227
          :
Db  181 A 181
```

RESULT 13
US-10-131-241-54
; Sequence 54, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.

Query Match   67.4%;   Score 835; DB 14;   Length 182;
Best Local Similarity   85.1%;   Pred. No. 2.7e-77;
Matches 154; Conservative 14; Mismatches 13; Indels   0; Gaps   0;

Qy   47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
          |::|||||:||||||:||||||||| ||||||||||||| |||||||||||||||||||||
Db   1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI   60

Qy 107 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 166
          ||||| :|:|||||||||||||||||||:||||||||||||||| ||:||:|||
Db   61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy 167 HGSDPSGRRLLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
          ||||| ||||||||||||||||||| |||||||||:||||||| |: ||||||||||||
Db 121 HGSDPNGRRLLTESYCETWRTEAPSATGQASSLLGQSAASCHHAYIVLCIENSFMT 180

Qy 227 S 227
          :
Db 181 A 181

RESULT 15
US-10-373-561-14
; Sequence 14, Application US/10373561
; Publication No. US20030175276A1
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/10/373,561
; CURRENT FILING DATE: 2003-02-24
; PRIOR APPLICATION NUMBER: US/09/561,499
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-10-373-561-14

Query Match   67.4%;   Score 835; DB 14;   Length 182;
Best Local Similarity   85.1%;   Pred. No. 2.7e-77;
Matches 154; Conservative 14; Mismatches 13; Indels   0; Gaps   0;

Qy   47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
          |::|||||:||||||:||||||||| ||||||||||||| |||||||||||||||||||||
Db   1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI   60

Qy 107 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 166
          ||||| :|:|||||||||||||||||||:||||||||||||||| ||:||:|||
Db   61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy 167 HGSDPSGRRLLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
          ||||| ||||||||||||||||||| |||||||||:||||||| |: ||||||||||||
Db 121 HGSDPNGRRLLTESYCETWRTEAPSATGQASSLLGQSAASCHHAYIVLCIENSFMT 180

Qy 227 S 227
          :
Db 181 A 181

Search completed: March 26, 2004, 13:47:33
Job time : 42.6667 secs

```
            GenCore version 5.1.6
   Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 13:33:07 ; Search time 15.5556 Seconds
                                    (without alignments)
                                    1422.260 Million cell updates/sec

Title:         US-09-938-391-2
Perfect score: 1239
Sequence:      1 PWRADDILAGPPRLLDPQPY.........CRHAFVVLCIENSVMTSFSK 230

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:      283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   PIR_78:*
             1: Pir1:*
             2: Pir2:*
             3: Pir3:*
             4: Pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Query Length	DB	ID	Description
1	1003	81.0	684	2	A53019	collagen alpha 1(X
2	983.5	79.4	1774	2	B56101	collagen alpha 1
3	980.5	79.1	1315	2	A56101	collagen alpha 1X
4	528.5	42.7	1388	2	A53317	collagen alpha 1X
5	369.5	29.8	650	2	T22002	hypothetical prote
6	99	8.0	244	2	C45729	sulfolipid biosynt
7	98	7.9	427	2	A53798	58K membrane-assoc
8	93.5	7.5	547	2	T44743	probable thiamin b
9	90	7.3	404	2	G90781	hypothetical prote
10	90	7.3	404	2	D85642	hypothetical prote
11	89	7.2	335	2	T34086	hypothetical prote
12	87.5	7.1	816	2	S46268	ataxin-1 - human
13	87	7.0	494	2	A83204	alginate biosynthe
14	85.5	6.9	508	2	T36945	hypothetical prote
15	85	6.9	300	2	G75295	hypothetical prote
16	85	6.9	312	2	A55461	NAD(P)-arginine AD
17	84	6.8	426	2	T17336	hypothetical prote
18	83	6.7	1187	1	TVHUY2	protein-tyrosine k
19	82.5	6.7	5232	2	A45086	HC-toxin synthetas
20	82	6.6	429	2	T20387	hypothetical prote
21	82	6.6	486	2	A41537	DNA-binding protei
22	82	6.6	995	2	H59432	RhoGAP protein hom
23	82	6.6	6420	2	T30283	polyketide synthas
24	81.5	6.6	190	2	T36950	probable transcrip
25	81.5	6.6	312	2	B55461	NAD(P)-arginine AD
26	81.5	6.6	566	2	T34842	probable transfera
27	81.5	6.6	633	2	G90704	probable Vgr prote
28	81.5	6.6	633	2	B85555	probable Vgr prote
29	81	6.5	833	2	T01547	probable phospholi
30	80.5	6.5	439	2	PQ0054	hypothetical prote
31	80	6.5	588	2	AC2276	oligopeptide bindi
32	80	6.5	624	2	B83386	hypothetical prote
33	80	6.5	1082	2	T45096	probable arabinosy
34	79.5	6.4	346	2	G83127	conserved hypothet
35	79.5	6.4	713	2	D90658	VgrG protein [impo
36	79.5	6.4	713	2	D85509	hypothetical prote
37	79	6.4	558	2	T30418	chitinase (EC 3.2.
38	79	6.4	672	2	T36083	hypothetical prote
39	79	6.4	1283	2	T49804	hypothetical prote
40	78.5	6.3	999	2	I38547	novel cellular pro
41	78.5	6.3	1049	2	T42045	beta transducin-li
42	78	6.3	345	2	C87288	hypothetical prote
43	78	6.3	479	2	AG2726	DNA photolyase [li
44	78	6.3	479	2	C97508	blue-light photore
45	78	6.3	511	2	S52622	amidophosphoribosy

ALIGNMENTS

```
RESULT 1
A53019
collagen alpha 1(XVIII) chain - human (fragment)
N;Contains: endostatin
C;Species: Homo sapiens (man)
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 15-Sep-2003
C;Accession: A53019
R;Oh, S.P.; Warman, M.L.; Seldin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olsen,
  Genomics 19, 494-499, 1994
A;Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localiz
A;Reference number: A53019; MUID:94245237; PMID:8188291
A;Accession: A53019
A;Molecule type: mRNA
A;Residues: 1-684 <OHA>
A;Cross-references: GB:L22548; NID:g348908; PIDN:AAA51864.1; PID:g562794
A;Note: the cited accession number, L25548, is not in Genbank release 103
A;Note: in the authors' translation, 482-Gly is not shown, residues 483-490 are shifted
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
lated and subsequently O-glycosylated.
C;Comment: Different splice forms of collagen alpha 1(XVIII) may be involved in perivas
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of u
ay be useful in treating solid tumors.
C;Genetics:
A;Gene: GDB:COL18A1
A;Cross-references: GDB:138752; OMIM:120328
A;Map position: 21q22.3-21q22.3
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteogly
F;1-684/Product: collagen alpha 1(XVIII) chain (fragment) #status predicted <MAT>
F;1-59/Domain: collagenous (fragment) #status predicted <CO4>
F;74-115/Domain: collagenous #status predicted <CO5>
F;129-201/Domain: collagenous #status predicted <CO6>
F;212-244/Domain: collagenous #status predicted <CO7>
F;257-278/Domain: collagenous #status predicted <CO8>
F;262-264/Region: cell attachment (R-G-D) motif
F;286-340/Domain: collagenous #status predicted <CO9>
F;354-371/Domain: collagenous #status predicted <CO10>
F;502-684/Domain: endostatin #status predicted <EST>
F;509-684/Region: multiplexin collagen carboxyl-terminal homologous

Query Match            81.0%;  Score 1003;  DB 2;  Length 684;
Best Local Similarity  81.5%;  Pred. No. 9.8e-84;
Matches  185;  Conservative  15;  Mismatches  23;  Indels  4;  Gaps  1;

Qy    1 PWRADDILAGPPRRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHTHQDFQLVHLVA 60
         |||||||||  |||||||||||||||||||||| |||||       |||:|||||||||||
Db  460 PWRADDILASPPGPLPEPQPYPGGPHHSSYVHCGPARPTSPP----AHSHRDFQPVLHLVA 515

Qy   61 LNSPQPGMRGIRGADFQCFCQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
         |||| :||||||||||||||||||| |||||||||||||||||||||||||| ||:||
Db  516 LNSPLSGGMRGIRGADFQCFCQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL 575

Qy  121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180
```

C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of u
ay be useful in treating solid tumors.
C;Genetics:
A;Gene: MGI:Col18a1
A;Cross-references: MGI:71175
A;Map position: 10:41.0
A;Introns: 195/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 159
A;Note: the list of introns is incomplete
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteogl
F;1-1774/Product: collagen alpha 1(XVIII) chain precursor, long splice form #status pre
F;-239,487-1774/Product: collagen alpha 1(XVIII) chain precursor, medium splice form #
F;1-24/Domain: signal sequence #status predicted <SIG>
F;786-812/Domain: collagenous #status predicted <CO01>
F;823-896/Domain: collagenous #status predicted <CO02>
F;921-1042/Domain: collagenous #status predicted <CO03>
F;1066-1148/Domain: collagenous #status predicted <CO04>
F;1163-1204/Domain: collagenous #status predicted <CO05>
F;1218-1290/Domain: collagenous #status predicted <CO06>
F;1301-1333/Domain: collagenous #status predicted <CO07>
F;1346-1369/Domain: collagenous #status predicted <CO08>
F;1351-1353/Region: cell attachment (R-G-D) motif
F;1377-1428/Domain: collagenous #status predicted <CO09>
F;1442-1459/Domain: collagenous #status predicted <CO10>
F;1591-1774/Product: endostatin #status predicted <EST>
F;1598-1774/Region: multiplexin collagen carboxyl-terminal homologous
P;354,361,947/Region: carbohydrate (Asn) (covalent) #status predicted
P;699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted
P;910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match   79.4%; Score 983.5; DB 2; Length 1774;
Best Local Similarity   80.1%; Pred. No. 1.9e-81;
Matches 185; Conservative 18; Mismatches 23; Indels 5; Gaps 2;

Qy   1 PWRADDILAGPRRLLDPQPYPGAP-HHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLV 59
           |||||||| | |||| | |||| |||| ||| |||| ||||| ||||
Db   1548 PWRADDILANPPRLLPDRQPYPGVPHHHSSYVHLLPPARPT----LSLAHTHQDFQPVLHLV 1603

Qy   60 ALNSPQPGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVN 119
           ||| | .|||||||||||||||||| .||.||||||||||||||||||||||| || |||
Db   1604 ALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIVN 1663

Qy   120 LRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDS 179
           |.||||.||| .|||||.||| |||||||||||||||.|||||.|||||||||||||.. :|
Db   1664 LKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLMES 1723

Qy   180 YCETWRTEAPAATGQASLLAGRLLEQRAASCRHAFVLCIENSVMTSFSK 230
           |||||||| ||||||||| ||||||| |||| .::||||||| |||||||
Db   1724 YCETWRTETTGATQQASLLSGRLLEQKAASCHNSYIVLCIENSFMTSFSK 1774

RESULT 3
A56101
collagen alpha 1(XVIII) chain precursor, short splice form - mouse
N;Contains: endostatin
C;Species: Mus musculus (house mouse)
C;Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text_change 15-Sep-2003
A;Accession: A56101; A58371; S72450; S65595
R;Rehn, M.; Pihlajaniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A;Title: Identification of three N-terminal ends of type XVIII collagen chains and tis
tif homologous to rat and Drosophila frizzled proteins.
A;Reference number: A56101; MUID:95181468; PMID:7876242
A;Accession: A56101
A;Molecule type: mRNA
A;Residues: 1-103 <REH1>
A;Cross-references: GB:U11636; NID:g618427; PIDN:AAC52178.1; PID:g618428
R;Rehn, M.; Pihlajaniemi, T.
Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994
A;Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collagenou
A;Reference number: A58371; MUID:94240112; PMID:8183894
A;Accession: A58371
A;Molecule type: mRNA
A;Residues: 1-928 <REH2>

Db   :||;||||||||||| |||||| |||||||||| ||||||||| ||:|||||||:||||:|
Db   576 KDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTESY 635

Qy   181 CETWRTEAPAATGQASLLAGRLLEQEAASCRHAFVVLCIENSVMTS 227
           ||||||||| |||||||| |||||| ||| .: ||||||||:
Db   636 CETWRTEAPSATGQASLLGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 682

RESULT 2
B56101
collagen alpha 1(XVIII) chain precursor, long splice form - mouse
N;Contains: collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin
C;Species: Mus musculus (house mouse)
C;Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text change 15-Sep-2003
A;Accession: B56101; C56101; S72450; S65595; PN0675; A54072; A58816
R;Rehn, M.; Pihlajaniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A;Title: Identification of three N-terminal ends of type XVIII collagen chains and tissu
tif homologous to rat and Drosophila frizzled proteins.
A;Reference number: A56101; MUID:95181468; PMID:7876242
A;Accession: B56101
A;Molecule type: mRNA
A;Residues: 1-562 <REH1>
A;Cross-references: GB:U11637; NID:g618429; PIDN:AAC52179.1; PID:g618430
A;Experimental source: splice form clone PE17.24
A;Accession: C56101
A;Molecule type: mRNA
A;Residues: 1-239,487-562 <REH2>
A;Cross-references: GB:U11637; NID:g618429
A;Accession: S72450
A;Molecule type: mRNA
A;Experimental source: splice form clones PE8.1, PE8.1, PE19, PE15.2
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A;Reference number: S72450
A;Accession: S72450
A;Molecule type: mRNA
A;Residues: 487-1146,'L',1148-1193,'F',1195-1210,'R',1212-1512,'L',1514-1522,'F',1524-16
A;Cross-references: EMBL:L22545; NID:g348968; PIDN:AAA19787.1; PID:g511298
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
A;Reference number: A58370; MUID:94240111; PMID:8183893
A;Accession: S65595
A;Molecule type: mRNA
A;Residues: 487-1512,'L',1514-1522,'F',1524-1683,'V',1685-1774 <OH2>
A;Cross-references: EMBL:L22545
R;Abe, N.; Muragaki, Y.; Yoshioka, H.; Inoue, H.; Ninomiya, Y.
Biochem. Biophys. Res. Commun. 196, 576-582, 1993
A;Title: Identification of a novel collagen chain represented by extensive interruptions
A;Reference number: PN0675; MUID:94059075; PMID:8240330
A;Accession: PN0675
A;Molecule type: mRNA
A;Residues: 635-1774 <ABE>
R;Rehn, M.; Hintikka, E.; Pihlajaniemi, T.
J. Biol. Chem. 269, 13929-13935, 1994
A;Title: Primary structure of the alpha1 chain of mouse type XVIII collagen, partial str
collagen chain.
A;Reference number: A54072; MUID:94245707; PMID:8188673
A;Accession: A54072
A;Molecule type: DNA; mRNA
A;Residues: 1293-1403,'R',1405-1774 <REH3>
A;Cross-references: GB:U03714; NID:g487733; PIDN:AAA20657.1; PID:g487734
R;O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukai, N.; Vasios, G.; Lane, W.S.; Flynn, E.; Bi
Cell 88, 277-285, 1997
A;Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.
A;Reference number: A58816; MUID:97160848; PMID:9008168
A;Accession: A58816
A;Molecule type: protein
A;Residues: 1591-1610 <ORE>
A;Experimental source: hemangioendothelium cells
A;Note: inhibits endothelial cell proliferation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
lated and subsequently O-glycosylated.
C;Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in peri

A;Cross-references: GB:L16898; NID:g404754; PIDN:AAA37434.1; PID:g553894
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A;Reference number: S72450
A;Molecule type: mRNA
A;Residues: 28-687,'L', 689-734,'F',736-751,'R',753-1315 <OHW>
A;Cross-references: EMBL:L22545; NID:g348968; PIDN:AAA19787.1; PID:g511298
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994.
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
A;Reference number: A58370; MUID:94240111; PMID:8183893
A;Accession: S65595
A;Molecule type: mRNA
A;Residues: 28-1315 <OHS>
A;Cross-references: EMBL:L22545
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
lated and subsequently O-glycosylated.
C;Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in peri
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un
ay be useful in treating solid tumors.
C;Genetics:
A;Gene: MGI:Col18a1
A;Cross-references: MGI:71175
A;Map position: 10:41.0
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc
F;1-25/Domain: signal sequence #status predicted <SIG>
F;24-235/Region: thrombospondin amino-terminal homologous
F;26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <M
F;327-353/Domain: collagenous #status predicted <CO1>
F;364-437/Domain: collagenous #status predicted <CO2>
F;462-583/Domain: collagenous #status predicted <CO3>
F;607-689/Domain: collagenous #status predicted <CO4>
F;704-745/Domain: collagenous #status predicted <CO5>
F;759-831/Domain: collagenous #status predicted <CO6>
F;842-874/Domain: collagenous #status predicted <CO7>
F;887-910/Domain: collagenous #status predicted <CO8>
F;892-894/Region: cell attachment (R-G-D) motif
F;918-969/Domain: collagenous #status predicted <CO9>
F;983-1000/Domain: collagenous #status predicted <CO10>
F;1132-1315/Product: collagen alpha 1(XVIII) chain #status predicted <EST>
F;1139-1315/Region: multiplexin collagen carboxyl-terminal homologous
F;1126,488/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;172-228/Disulfide bonds: #status predicted
F;240,245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;451,454,594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match   79.1%; Score 980.5; DB 2; Length 1315;
Best Local Similarity 79.7%; Pred. No. 2.5e-81;
Matches 184; Conservative 19; Mismatches 23; Indels 5; Gaps 2;

Qy   1 PWRADDILAGPPRLLDPQYPGAP-HHGSYVHFQPARPTGGPVHTHTHTHQDFQLVLHLV 59
            |||||||||| ||| || :::|||||||||||||||||||||||||||||||
Db   1089 PWRADDILANPPRLPDRQPYPGVPHHHSSVHLPPARPT----LSLAHTHQDFQPVLHLV 1144

Qy   60 ALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVN 119
            ||| ||.||||||||||||||||||| |||||||||||||||||||||||||: ||
Db   1145 ALNTPLSGGWMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIVN 1204

Qy   120 LRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDS 179
            :|::|||||| ||||:||||||||||||||||||:|||:|||:||||||||||||| :
Db   1205 LKDEVLSPSWNDSLFSGSQGQVQPGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLMES 1264

Qy   180 YCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 230
            ||||||| ||.||||||||||||||||||||:||||||||||||||||||
Db   1265 YCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMTSFSK 1315

RESULT 4
A53317
collagen alpha 1(XV) chain precursor - human
N;Alternate names: procollagen alpha 1(XV) chain
C;Species: Homo sapiens (man)

C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 15-Sep-2003
C;Accession: A53317; A53146; S28778
R;Kivirikko, S.; Heinaemaeki, P.; Rehn, M.; Honkanen, N.; Myers, J.C.; Pihlajaniemi, T.
J. Biol. Chem. 269, 4773-4779, 1994
A;Title: Primary structure of the alpha1 chain of human type XV collagen and exon-intrc
A;Reference number: A53317; MUID:94148920; PMID:8106446
A;Accession: A53317
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1388 <KIV>
A;Cross-references: GB:L25280
A;Note: nucleotide sequence and conceptual translation not complete
R;Muragaki, Y.; Abe, N.; Ninomiya, Y.; Olsen, B.R.; Ooshima, A.
J. Biol. Chem. 269, 4042-4046, 1994
A;Title: The human alpha1(XV) collagen chain contains a large amino-terminal non-triple
A;Reference number: A53146; MUID:94140817; PMID:8307960
A;Accession: A53146
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-9,'S',11-48,'V',50-94,'A',96-149,'A',151-203,'V',205-408,'A',410-569 <MUR
A;Cross-references: GB:D21230; NID:g415605; PIDN:BAA04762.1; PID:d1005294; PID:g460703
R;Myers, J.C.; Kivirikko, S.; Gordon, M.K.; Dion, A.S.; Pihlajaniemi, T.
Proc. Natl. Acad. Sci. U.S.A. 89, 10144-10148, 1992
A;Title: Identification of a previously unknown human collagen chain, alpha1(XV), chara
A;Reference number: S28778; MUID:93066196; PMID:1279671
A;Accession: S28778
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 544-640,'P',642-811,'P',813-1252 <MYB>
C;Genetics:
A;Gene: GDB:COL15A1
A;Cross-references: GDB:132578; OMIM:120325
A;Map position: 9q21-9q22
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-1388/Product: collagen alpha 1(XV) chain #status predicted <MAT>
F;1216-1388/Region: multiplexin collagen carboxyl-terminal homologous

Query Match   42.7%; Score 528.5; DB 2; Length 1388;
Best Local Similarity 48.2%; Pred. No. 5.8e-40;
Matches 106; Conservative 30; Mismatches 63; Indels 63; Gaps 3;

Qy   11 PPRLLDPQYPGAPHHGSYVHFQPARPTGGPVHTHTHTHQDFQLVLHVALNSPQPGGMR 70
            :|: ::||: |:||:||: ||||||||||||||||:||| ||: ||||||
Db   1190 PPPALSSNPHQLLP--------PPNPISSANYE------KPALHLAALNMPFSGDIR 1232

Qy   71 GIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNLRDEVLFPSWE 130
            |||||||| |||||||| ||| |||||:||||||| :|| ::|.|:||.||.:: :
Db   1233 ----ADFQCFKQARAAGLLSTIVRKAERYSLPIVNLKGGVLFNNWD 1288

Qy   131 ALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSYCETWRTEAPA 190
            :||||:||:||:||||||||||||:|||:|||: ||||| ||| ||:|:|||||||: .:
Db   1289 SIFSGHGGQFNMHIPIYSFDGRDIMTDSWPQKVIWHGSSPHGVRLVDNYCEAWRTADTA 1348

Qy   191 ATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 230
            .||:||| | ||:||: ::|||:|:||||||||||:|. |
Db   1349 VTGLASPLSTGKILDQKAYSCANRLIVLCIENSFMTDARK 1388

RESULT 5
T22002
hypothetical protein F39H11.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
A;Accession: T22002
R;White, S.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19500
A;Accession: T22002
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-650 <WIL>
A;Cross-references: EMBL:Z81079; PIDN:CAB03084.1; GSPDB:GN00019; CESP:F39H11.4
A;Experimental source: clone F39H11

```
R;Juang, S.H.; Huang, J.; Li, Y.; Salas, P.J.I.; Fregien, N.; Carraway, C.A.C.; Carraway
J. Biol. Chem. 269, 15067-15075, 1994
A;Title: Molecular cloning and sequencing of a 58-kDa membrane- and microfilament-assoc
A;Reference number: A53798; MUID:94253065; PMID:8195143
A;Accession: A53798
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-427 <JUA>
A;Cross-references: GB:U15425; GB:S70214; NID:G562083; PIDN:AAB54085.1; PID:g562084
C;Superfamily: mammalian retrovirus gag polyprotein I
C;Keywords: actin binding; monomer; phosphoprotein

Query Match       7.9%;    Score 98; DB 2;   Length 427;
Best Local Similarity    24.9%;    Pred. No. 0.32;
Matches    59;  Conservative    29;  Mismatches    97;  Indels    52;  Gaps    11;

Qy     5  DDILAGPPRLLDPQPYPGAPHHGSVHFQPARPTGGPVHTHTHQDFQLVLHLVALNSP     64
Db   156  DLLLEEP----PYPVPTAPPREEVE-PPARP----------------RLEAAPSP     191

Qy    65  QPGGMRGIR----GADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRR---ADRTGVPV   117
Db   192  VAGRLRGRREVAPDSTSQAFPLRQGAGGQMQYWPFSAA---DIYNWKQHNPPFSKDPVAL   248

Qy   118  VNLRDEVLF---PSW-------EALFSGSEGQ------LKPGARIFSFDGRDVLQ---HPAW  160
Db   249  TNLIESVLLTHQPTWDDIQQLQALLTSBEKQRVLLEARKHVLGDNGRPTLLPEEIDDAF     308

Qy   161  P-RKSVWHGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFV      216
Db   309  PLTRPDWDFFTAEGRRHLRLYRQLLLAGLRGAARRPTNLAQVKQVVQEAAETPSAFL      365

RESULT 8
T44743
probable thiamin biosythesis protein thiC [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Feb-2000
C;Accession: T44743
R;James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, January 1998
A;Reference number: Z22831
A;Accession: T44743
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-547 <JAM>
A;Cross-references: EMBL:AL035159; PIDN:CAA22712.1
A;Experimental source: cosmid B1450
C;Genetics:
A;Gene: thiC
C;Superfamily: thiamin biosynthesis protein thiC

Query Match       7.5%;    Score 93.5; DB 2;   Length 547;
Best Local Similarity    27.4%;    Pred. No. 1.1;
Matches    43;  Conservative    14;  Mismatches    49;  Indels    51;  Gaps    8;

Qy    19  PYPGAPHHGSVHFQPARPTGGP---------VHTHTHTHQDF-------QLVLHLVA    60
Db    17  PITPGS---SKAYREVANPDGGPSLRVFPRRVHLSTGAHFDLYDTSGPYTDPDAVINLTA  72

Qy    61  LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL  120
Db    73  GLPPRPGVIRD-RGTQ--LQRARAGEITAEM-AFI----------ADREGMPAELV      114

Qy   121  RDEVLF-----------PSWEALFSGSEGQLKPGARI    146
Db   115  RVEVALGRAVIPANHNHPEIEPMIIGKAFAVKVNANI    151

RESULT 9
G90781
hypothetical protein ECs1223 [imported] - Escherichia coli (strain O157:H7, substrain I
C;Species: Escherichia coli
```

```
C;Genetics:
A;Gene: CESP:F39H11.4
A;Map position: 1
A;Introns: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

Query Match      29.8%;    Score 369.5; DB 2;   Length 650;
Best Local Similarity    41.7%;    Pred. No. 8e-26;
Matches    80;  Conservative    27;  Mismatches    70;  Indels    15;  Gaps    6;

Qy    33  QPARPTGGPVHTHTHQDFQLVLHLVALNSPQGGMRGIRGADFQCFQQARAAGLAGTF     92
Db   451  EPQRDAG--VHNKDR------VIHMIRLSQPFSGNLHGLRGADLQCYREARAAGYTTTF    501

Qy    93  RAFLSRLQDLYSIVRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGR    152
Db   502  RAMLSSNVQDLVRIVHSVD-FDTTVVNVAGHHLFPSWRSFVNGA--QMNPHAXLFSFDRH  558

Qy   153  DVLQHPAWPRKSVWHGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAAS--  210
Db   559  DVLNDSRWPDKRVWHGSKDGGIR-AEQYCDGWRRADSSLTSLAGHISSNTSIFQSSGSEK  617

Qy   211  CRHAFVVLCIEN    222
Db   618  CENKLVVLCVEN    629

RESULT 6
C45729
sulfolipid biosynthesis protein sqdC - Rhodobacter sphaeroides
C;Species: Rhodobacter sphaeroides
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999
C;Accession: C45729; S27664
R;Benning, C.; Somerville, C.R.
J. Bacteriol. 174, 6479-6487, 1992
A;Title: Identification of an operon involved in sulfolipid biosynthesis in Rhodobacter
A;Reference number: A45729; MUID:93015699; PMID:1400200
A;Accession: C45729
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-244 <BEN>
A;Cross-references: GB:M89780; NID:g152038; PIDN:AAA73225.1; PID:g152041
C;Genetics:
A;Gene: sqdC
A;Start codon: GTG

Query Match       8.0%;    Score 99; DB 2;   Length 244;
Best Local Similarity    23.9%;    Pred. No. 0.13;
Matches    56;  Conservative    28;  Mismatches    96;  Indels    54;  Gaps    8;

Qy    14  LLDPQPYPGAPHHGSVHFQPARPTGGPVHTHTHQDFQLVLHLVALNSPQPGGMRGIR    73
Db     2  ILARSPLPGADEHRPYDLLGPLPLA---------DVDALIHCAFQH--VPGRYRGGE    48

Qy    74  GADFQCFQQARAAGLAGTFRAFLSRLQDLYSIVRRADRTGVPVVNLRDEVLFPSWEALF   133
Db    49  GQDPBGF---RRANLEGSLRLFEAMRGRRILFLSSRAVFDGYGPGTLLTEAMPPCPESLY  105

Qy   134  SGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSYCETWR-------  185
Db   106  ---GQVKABAEVALF-----AEGGASLRATGVYGPGPDHK--------WRVLFEDFRA   147

Qy   186  --------TEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK    230
Db   148  GRPIEPRVATEVHGADLAAAALL--LLEKPDAGAFHVSDLLLDRHDLLAEVAR    198

RESULT 7
A53798
58K membrane-associated protein - rat
N;Alternate names: 58K microfilament-associated protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 24-Nov-1999
C;Accession: A53798
```

C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: G90781
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: G90781
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-404 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB34646.1; PID:g13360683; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: BCs1223

Query Match   7.3%;   Score 90; DB 2; Length 404;
Best Local Similarity 25.5%;   Pred. No. 1.6;
Matches   73; Conservative   28; Mismatches 107; Indels   78; Gaps   18;

Qy   9 AGPP--RLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTH----THQDFQLVL----HL 58
           :||| ||:|:   :||||||||||||
Db   53 AGAPVVRITDLNKQAGDEVTFSIMHKLSKRPTMGDERVEGRGEDLSHADFSLKINQGRHL 112

Qy   59 VALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRR-------A 110
           | :|:   ::| |
Db   113 V-----DAGGRMSQQRTKFNLASSART--LLGTY----FNDLQDQCAIVHLAGARGDFVA 161

Qy   111 DRTGVPVVN-------LRDEVLFPSWEALFSGSBG----QLKPGARIFSF---DGR---- 152
              :||| :|:|: | ::|
Db   162 DDTILPTAEHPEFKKIMINDVLPPTHDRHFFGGDATSFEQIE-AADIFSIGLVDNLSLFI 220

Qy   153 DVLQHPAWPRK---SVWHGSDP------SGRRLTDSYCET----WRTEAPAATGQASSLL 199
              : :|| |||| ||||:|: |
Db   221 DEMAHPLQPVRLSGDELHGEDPYYVLYVTPRQWNDWYTSTSGKDWNQMMVRAVNRAKG-F 279

Qy   200 AGRLLEQBAASCRHAFV-------------VLCIENSVMTSFSK 230
             |:|||::|| |:   :|||||||||||
Db   280 NHPLFKGECAMWRNILVRKYAGMPIRFYQGSKVLVSENN-LTATTK 324

RESULT 10
D85642
hypothetical protein Z1479 (imported) - Escherichia coli (strain O157:H7, substrain EDL9
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: D85642
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D85642
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-404 <STO>
A;Cross-references: GB:AE005174; NID:g12514332; PIDN:AAG55600.1; GSPDB:GN00145; UWGP:Z14
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z1479

Query Match   7.3%;   Score 90; DB 2; Length 404;
Best Local Similarity 25.5%;   Pred. No. 1.6;
Matches   73; Conservative   28; Mismatches 107; Indels   78; Gaps   18;

Qy   9 AGPP--RLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTH---THQDFQLVL----HL 58
           :||| ||:|:   :||||||||||||
Db   53 AGAPVVRITDLNKQAGDEVTFSIMHKLSKRPTMGDERVEGRGEDLSHADFSLKINQGRHL 112

Qy   59 VALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRR-------A 110
           | :|:   ::| |
Db   113 V-----DAGGRMSQQRTKFNLASSART--LLGTY----FNDLQDQCAIVHLAGARGDFVA 161

Qy   111 DRTGVPVVN-------LRDEVLFPSWEALFSGSBG----QLKPGARIFSF---DGR---- 152
              :||| :|:|: | ::|
Db   162 DDTILPTAEHPEFKKIMINDVLPPTHDRHFFGGDATSFEQIE-AADIFSIGLVDNLSLFI 220

Qy   153 DVLQHPAWPRK---SVWHGSDP------SGRRLTDSYCET----WRTEAPAATGQASSLL 199
              : :|| |||| ||||:|: |
Db   221 DEMAHPLQPVRLSGDELHGEDPYYVLYVTPRQWNDWYTSTSGKDWNQMMVRAVNRAKG-F 279

Qy   200 AGRLLEQBAASCRHAFV-------------VLCIENSVMTSFSK 230
             |:|||::|| |:   :|||||||||||
Db   280 NHPLFKGECAMWRNILVRKYAGMPIRFYQGSKVLVSENN-LTATTK 324

RESULT 11
T34086
hypothetical protein C06E7.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34086
R;Le, T.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid C06E7.
A;Reference number: Z21474
A;Accession: T34086
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-335 <LET>
A;Cross-references: EMBL:U41009; PIDN:AAA82281.1; CESP:C06E7.4
C;Genetics:
A;Gene: CESP:C06E7.4
A;Introns: 41/2; 83/1; 116/3; 219/1; 273/3

Query Match   7.2%;   Score 89; DB 2; Length 335;
Best Local Similarity 30.3%;   Pred. No. 1.6;
Matches   20; Conservative   4; Mismatches   22; Indels   20; Gaps   3;

Qy   1 PWRADDILAGPPRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLVA 60
            | :: |||||
Db   71 PWNTSSISSGKRDIISPPMYSSTPHH------------------HNH-HNHQQLQ---HHQV 110

Qy   61 LNSPQP 66
            ||||
Db   111 IKSPQP 116

RESULT 12
S46268
ataxin-1 - human
C;Species: Homo sapiens (man)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999
C;Accession: S46268
R;Banfi, S.; Servadio, A.; Chung, M.; Kwiatkowski Jr., T.J.; McCall, A.E.; Duvick, L.A
Nature Genet. 7, 513-519, 1994
A;Title: Identification and characterization of the gene causing type 1 spinocerebellar
A;Reference number: S46268; MUID:95038838; PMID:7951322
A;Accession: S46268
A;Molecule type: mRNA
A;Residues: 1-816 <BAN>
A;Cross-references: EMBL:X79204; NID:g529661; PIDN:CAA55793.1; PID:g529662
C;Comment: Spinocerebellar ataxia type 1, a dominantly inherited neurodegenerative diso
repeats in SCA1 patients.
C;Genetics:
A;Gene: GDB:SCA1
A;Cross-references: GDB:119588; OMIM:164400
A;Map position: 6p23-6p23

Query Match   7.1%;   Score 87.5; DB 2; Length 816;
Best Local Similarity 22.4%;   Pred. No. 6.5;
Matches   54; Conservative   26; Mismatches   96; Indels   65; Gaps   10;

Qy   8 LAGPPRLLDPQPYPGAPHHGSYVHF-------QPARPTGGPVHTHTHQDFQLVLHLHV 59
           || || :| |
Db   228 LSRAPGLIITPGSPPPA-QQNQVHISSSPQNTGRIASPPAIPVHLHPHQ----TMIPHTL 282

```
Qy   60 ALNSPQPGGMR-GIRGADFQCFQQARAAGLAGTFRAFLSSRLQD------LYSIVRRAD 111
           |||  | |: |:| ||| ||:   :                      |:||||||
Db  283 TLGPPSQVVMQYADSGSHFVPREATKKAE---------SSRLQQAIQAKEVLNGEMEKSR 333

Qy  112 RTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARI--FSFDGRDVLQHPAWPRKSVWHGS 169
           | |:           ||| | :| :   |  :|    |  |
Db  334 RYGAP------------------SSADLGLGKAGGKSVPHPYESRHVVVHPS--PSDYSSR 374

Qy  170 DPSGRRLT---------------DSYCETWRTEAPAATGQASLLAGRLLEQEAASCRHAF 215
          |||| |:                ||   |::|| :  :| :
Db  375 DPSGVRASVMVLPNSNTPAADLEVQQATHREASPSTLNDKSGLHLGKPGHRSYALSPHTV 434

Qy  216 V 216
          :
Db  435 I 435


RESULT 13
A83204
alginate biosynthesis protein Alg8 PA3541 [imported] - Pseudomonas aeruginosa (strain PA
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83204
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
  adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
  .; Lory, S.; Olson, M.V.
  Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A83204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-494 <STO>
A;Cross-references: GB:AE004774; GB:AE004091; NID:g9949684; PIDN:AAG06929.1; GSPDB:GN001
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA3541

Query Match        7.0%;    Score 87;  DB 2;  Length 494;
Best Local Similarity  24.6%;    Pred. No. 3.9;
Matches   45;  Conservative  20;  Mismatches  68;  Indels  50;  Gaps  7;

Qy   50 QDFQLVLHLVALNSPQPGGMRGIRGADF----------QCFQQARAAGLAGTFRAFLSSR  99
          :||:|::| |: | | :  | :| :|           :|: : |    | :  :: | |
Db   39 KDFILLIGAVGINRYSMGGVHFLRGMLFLHVYYPYYRRRVRQLGSAADPSHVFLMVTSFR  98

Qy  100 LQDL------YSIVRRADRTGVP------VVNLRDEVLFPS-WEAL--------------  132
          : |       |::|  | :|        :| :  :|:|   ||
Db   99 IDALTTAMVYRSVIREAIDSGYPTTVVCSIVEMSDEVLVRSLWEKMNPPDRVSLDFVRIP 158

Qy  133 FSGSEGQLKPGARIFS----------FDGRDVLQHPAWPRKSVWHGSDPS-GRRLTDS  179
           :|   |   |  ::            |  | :  |          | :  |:: |:
Db  159 GTGKRDGLAYGFRAISRHLPDDDAVVAVIDGDTVLDHGVVKKTVPWFKLFPNVGGLITNE 218

Qy  180 YCE 182
          :|
Db  219 FCE 221


RESULT 14
T36945
hypothetical protein SCJ1.12 - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T36945
R;Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
  submitted to the EMBL Data Library, September 1999
A;Reference number: Z21607
A;Accession: T36945
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-508 <SEE>
```

```
A;Cross-references: EMBL:AL109962; PIDN:CAB53130.1; GSPDB:GN00070; SCOEDB:SCJ1.12
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCJ1.12

Query Match        6.9%;    Score 85.5;  DB 2;  Length 508;
Best Local Similarity  22.6%;    Pred. No. 5.6;
Matches   60;  Conservative  30;  Mismatches 100;  Indels  75;  Gaps  13;

Qy    4 ADDI--LAGPPRLLDPQPYP----------------------GAPHGSYVHFQPARPTG  39
          ||:||  | |  :|                             ||:|||:|||  ||
Db  204 ADDVFCLGDGPRILDCLEFDDALRYVDGLDDAAFLAMDLESLGAPESAAFFLAQYGBYSG  263

Qy   40 GPV------HTHTHTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARA-------A  86
           |||      :|                       || ||: |:|           |
Db  264 DPAPPSLWHHYVAYRAPFVRAKVSLIQARCGAPGAHATARRLVRMALRHLRASAVGLTLVA 323

Qy   87 GLAGTFRAFLSSRLQD-LYSIVRRADRTGVPVVNLRDEV--LFPSWEALFSGSEGQLKP- 142
          ||||||||||||||||  |||||||                :|:||||||||||| :|
Db  324 GLPGTGKSTLSGALADRLGAVLLSSDR-------LRKEMAGLSPQQTASADYGEGLYTPE 376

Qy  143 -GARIFS--FD------GRDVLQHPAW----PRKSVWHGSDPSGRRLTDSYCET---- 183
           ||:|:   |          ||:||||     |:|||||||||| |||  :|||
Db  377 WTARTYAELLDRAAALLALGESVVLDATWIDSAQREARHTAESAGADLVALHCHVPDDV 436

Qy  184 ----WRTEAPAAT---GQASLLA  200
            :|||:|:        |||::|
Db  437 TAARLSTRAPGASDADLGVARAMAA 461


RESULT 15
G75295
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: G75295
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
  M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; t
  S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
  Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: G75295
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-300 <WHI>
A;Cross-references: GB:AE002058; GB:AE002059; NID:g6460059; PIDN:AAF11808.1; PID:g6460(
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2255
A;Map position: 1

Query Match        6.9%;    Score 85;  DB 2;  Length 300;
Best Local Similarity  22.1%;    Pred. No. 3.2;
Matches   54;  Conservative  17;  Mismatches  87;  Indels  86;  Gaps  10;

Qy   23 APHHGSYV----------------HFQPARPTGGPVHTHTHQDFQLVLHLVALNSPQP  66
          ||:||:|                 :|  |  || |:||:|:   | :|:||   |
Db   45 APEHGYVVSETAQGVSGLGSLWLPDFHPSHAWG-LHLHPDHHADAAALLEFLAAQARAA 103

Qy   67 G------------------GMR------GIRGADFQCFQQARAAGL--------------  88
          |                  | :      ||  | :   |
Db  104 GRSHLWASVREDYLPAWEDLPALGFREVHRTGGGFFLRDWQVNTGALEALTARRYRLT 163

Qy   89 -AGTFRAFLSSRLQDLYSIVRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIF 147
           |   | :   | :||:  :    | :          |   |
Db  164 PAAFPQH--DARLTELYALTRQDVTAPPTID------PAAEAL--------------- 199

Qy  148 SFDGRDVLQHPAWPRKSVWHGSDPSGRRLTD-SYCETWRTEAPAATGQASLLAGRLLEQ 206
           ||| | :  |          |:  : |: :  : | : |  |   |  :|| :| |
Db  200 --TDEDALWNAAN--LAWHGETLVGLALPERSRLHAWNAVLVVHPEHRRQGLATALLAQ 254

Qy  207 EAAS 210
```

us-09-938-391-2.rpr

Db   255 VARS 258

Search completed: March 26, 2004, 13:38:38
Job time : 19.5556 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2004, 13:29:16 ; Search time 11.6667 Seconds
(without alignments)
1026.526 Million cell updates/sec

Title: US-09-938-391-2
Perfect score: 1239
Sequence: 1 PWRADDILAGPPRLLDPQPY.........CRHAFVVLCIENSVMTSFSK 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1019	82.2	1516	1	CA1H_HUMAN	P39060 homo sapien
2	983.5	79.4	1774	1	CA1H_MOUSE	P39061 mus musculu
3	528.5	42.7	1388	1	CA1E_HUMAN	P39059 homo sapien
4	93.5	7.5	547	1	THIC_MYCLE	Q9zbl0 mycobacteri
5	90	7.3	556	1	GLI_CHICK	P55878 gallus gall
6	88	7.1	1233	1	NME3_HUMAN	Q14957 homo sapien
7	87.5	7.1	816	1	ATX1_HUMAN	P54253 homo sapien
8	87	7.0	494	1	ALG8_PSEAE	Q52463 pseudomonas
9	86	6.9	448	1	HGD_BRAJA	Q89xh1 bradyrhizob
10	85	6.9	312	1	NRTI_CHICK	P55806 gallus gall
11	84	6.8	948	1	CHRD_MOUSE	Q92oe2 mus musculu
12	82.5	6.7	5217	1	HTS1_COCCA	Q01886 cochliobolu
13	82	6.6	486	1	EGR4_HUMAN	Q05215 homo sapien
14	82	6.6	995	1	SR13_HUMAN	Q9y1m8 homo sapien
15	81.5	6.6	312	1	NRT2_CHICK	P55807 gallus gall
16	80.5	6.5	439	1	YT55_STRFR	P20190 streptomyce
17	80.5	6.5	7073	1	R1AB_CVHSA	P59641 h replicase
18	80	6.5	1082	1	EMBB_MYCSM	Q50395 mycobacteri
19	78.5	6.3	999	1	MERK_HUMAN	Q12866 homo sapien
20	78	6.3	1616	1	APXL_HUMAN	Q13796 homo sapien
21	77	6.2	248	1	PSPA_RAT	P08427 rattus norv
22	77	6.2	419	1	P47K_PSECL	P31521 pseudomonas
23	77	6.2	915	1	YQO7_CAEEL	Q09446 caenorhabdi
24	77	6.2	1187	1	TYK2_HUMAN	P29597 homo sapien
25	76.5	6.2	953	1	CAR4_MOUSE	Q8bhb0 mus musculu
26	76	6.1	724	1	MALQ_MYCTU	O53932 mycobacteri
27	76	6.1	726	1	PATA_VIBAN	P11461 vibrio angu
28	75.5	6.1	512	1	PPX_ECOLI	P29014 escherichia
29	75.5	6.1	838	1	GLGB_STRAW	Q82f60 streptomyce
30	75.5	6.1	1855	1	MY5A_HUMAN	Q9y41l homo sapien
31	75	6.1	300	1	NARE_CHICK	Q92080 gallus gall
32	75	6.1	303	1	Y247_HUMAN	Q92537 homo sapien
33	75	6.1	387	1	TIZ2_MOUSE	Q9eqn3 mus musculu
34			75	1	RECA_MYCTH	Q9f407 mycobacteri
35			75	1	IBMP_CAMVS	P03559 cauliflower
36		6.1	74.5	1	MURG_YERPE	Q8zie9 yersinia pe
37		6.1	74.5	1	AMRP_HUMAN	P30533 homo sapien
38		6.0	74.5	1	LOXX_SOYBN	P24095 glycine max
39		6.0	74.5	1	GLI1_HUMAN	P08151 homo sapien
40		6.0	74.5	1	IRS2_HUMAN	Q9y4h2 homo sapien
41		6.0	74	1	GLMS_MYCSM	O68956 m glucosami
42		6.0	74	1	TEGU_HSV11	P10220 herpes simp
43		6.0	73.5	1	HMD1_CHICK	P46692 gallus gall
44		5.9	73.5	1	PANC_YERPE	Q8zbk7 yersinia pe
45		5.9	73.5	1	PIM3_HUMAN	Q86v86 homo sapien

ALIGNMENTS

RESULT 1
CA1H_HUMAN
ID   CA1H_HUMAN   STANDARD;   PRT;   1516 AA.
AC   P39060; Q9UK38; Q9Y6Q7; Q9Y6Q8;
DT   01-FEB-1995 (Rel. 31, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin].
GN   COL18A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9816096; PubMed=9503365;
RA   Saarela J., Ylikarpa R., Rehn M., Purmonen S., Pihlajaniemi T.;
RT   "Complete primary structure of two variant forms of human type XVIII
RT   collagen and tissue-specific differences in the expression of the
RT   corresponding transcripts.";
RL   Matrix Biol. 16:319-328(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20289799; PubMed=10830953;
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA   Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
RA   Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA   Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
RN   [3]
RP   SEQUENCE OF 834-1516 FROM N.A.
RX   MEDLINE=94245237; PubMed=8182291;
RA   Oh S.P., Warman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,
RA   Olsen B.R.;
RT   "Cloning of cDNA and genomic DNA encoding human type XVIII collagen
RT   and localization of the alpha 1(XVIII) collagen gene to mouse
RT   chromosome 10 and human chromosome 21.";
RL   Genomics 19:494-499(1994).
RN   [4]
RP   SEQUENCE OF 1334-1516 FROM N.A.
RC   TISSUE=Placenta;
RA   Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.;
RT   "Cloning and expression of human endostatin gene in Escherichia
RT   coli.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   INVOLVEMENT IN KNOBLOCH SYNDROME.

```
RX   MEDLINE=20400145; PubMed=10942434;
RA   Sertie A.L., Sossi V., Camargo A.A., Zatz M., Brahe C.,
RA   Passos-Bueno M.R.;
RT   "Collagen XVIII, containing an endogenous inhibitor of angiogenesis
RT   and tumor growth, plays a critical role in the maintenance of retinal
RT   structure and in neural tube closure.";
RL   Hum. Mol. Genet. 9:2051-2058(2000).
RN   [6]
RP   VARIANT ASN-1437.
RX   MEDLINE=21518361; PubMed=11606364;
RA   Iughetti P., Suzuki O., Godoi P.H., Alves V.A., Sertie A.L.,
RA   Zorick T., Soares F., Camargo A., Moreira E.S., di Loreto C.,
RA   Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;
RT   "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes
RT   for the development of prostatic adenocarcinoma.";
RL   Cancer Res. 61:7375-7378(2001).
CC   -!- FUNCTION: COLA18A probably plays a major role in determining the
CC       retinal structure as well as in the closure of the neural tube.
CC   -!- FUNCTION: Endostatin potently inhibits endothelial cell
CC       proliferation and angiogenesis. May inhibit angiogenesis by
CC       binding to the heparan sulphate proteoglycans involved in growth
CC       factor signalling.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long; Synonyms=NC-493;
CC         IsoId=P39060-1; Sequence=Displayed;
CC       Name=Short; Synonyms=NC1-303;
CC         IsoId=P39060-2; Sequence=VSP_001155, VSP_001156;
CC   -!- TISSUE SPECIFICITY: Present in multiple organs with highest levels
CC       in liver, lung and kidney.
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC       unit (G-X-Y) are hydroxylated in some or all of the chains.
CC   -!- POLYMORPHISM: There is an association between a polymorphism in
CC       position 1437 and prostate cancer. Heterozygous Asn-1437
CC       individuals have a 2.5 times increased chance of developing
CC       prostate cancer as compared with homozygous Asp-1437 individuals.
CC   -!- DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KNO)
CC       [MIM:267750], an autosomal recessive disorder defined by the
CC       occurrence of high myopia, vitreoretinal degeneration with retinal
CC       detachment, macular abnormalities and occipital encephalocele.
CC   -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC       INTERRUPTED HELICES (FACIT) FAMILY.
CC   -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AF018081; AAC39658.1; -.
DR   EMBL; AF018082; AAC39659.1; -.
DR   EMBL; AL163302; CAB90482.1; -.
DR   EMBL; L22548; AAA51864.1; -.
DR   EMBL; AF184060; AAF01310.1; ALT_INIT.
DR   PDB; 1BNL; 02-DEC-98.
DR   GlycoSuiteDB; P39060; -.
DR   Genew; HGNC:2195; COL18A1.
DR   MIM; 120328; -.
DR   MIM; 267750; -.
DR   GO; GO:0005581; C:collagen; TAS.
DR   GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR   GO; GO:0008285; P:negative regulation of cell proliferation; TAS.
DR   GO; GO:0007601; P:vision; TAS.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 7.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
KW   Polymorphism; 3D-structure.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24   1516       COLLAGEN ALPHA 1 (XVIII) CHAIN.
FT   CHAIN      1334   1516       ENDOSTATIN.
FT   DOMAIN      221    409       TSP N-TERMINAL.
FT   DOMAIN      410    516       NONHELICAL REGION 1 (NC1).
FT   DOMAIN      517    550       TRIPLE-HELICAL REGION 1 (COL1).
FT   DOMAIN      551    560       NONHELICAL REGION 2 (NC2).
FT   DOMAIN      561    640       TRIPLE-HELICAL REGION 2 (COL2).
FT   DOMAIN      641    664       NONHELICAL REGION 3 (NC3).
FT   DOMAIN      665    786       TRIPLE-HELICAL REGION 3 (COL3).
FT   DOMAIN      787    809       NONHELICAL REGION 4 (NC4).
FT   DOMAIN      810    892       TRIPLE-HELICAL REGION 4 (COL4).
FT   DOMAIN      893    906       NONHELICAL REGION 5 (NC5).
FT   DOMAIN      907    948       TRIPLE-HELICAL REGION 5 (COL5).
FT   DOMAIN      949    961       NONHELICAL REGION 6 (NC6).
FT   DOMAIN      962   1034       TRIPLE-HELICAL REGION 6 (COL6).
FT   DOMAIN     1035   1044       NONHELICAL REGION 7 (NC7).
FT   DOMAIN     1045   1077       TRIPLE-HELICAL REGION 7 (COL7).
FT   DOMAIN     1078   1089       NONHELICAL REGION 8 (NC8).
FT   DOMAIN     1090   1111       TRIPLE-HELICAL REGION 8 (COL8).
FT   DOMAIN     1112   1118       NONHELICAL REGION 9 (NC9).
FT   DOMAIN     1119   1173       TRIPLE-HELICAL REGION 9 (COL9).
FT   DOMAIN     1174   1186       NONHELICAL REGION 10 (NC10).
FT   DOMAIN     1187   1204       TRIPLE-HELICAL REGION 10 (COL10).
FT   DOMAIN     1205   1516       NONHELICAL REGION 11 (NC11).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    129    129       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    164    164       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    691    691       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1329   1329       O-LINKED (GALNAC...).
FT                                /FTId=CAR_000150.
FT   DISULFID   1366   1506       BY SIMILARITY.
FT   DISULFID   1468   1498       BY SIMILARITY.
FT   SITE       1095   1097       CELL ATTACHMENT SITE (POTENTIAL).
FT   VARSPLIC      1    180       Missing (in isoform Short).
FT                                /FTId=VSP_001155.
FT   VARSPLIC    181    215       HTTEAGTLPAPTPSPPSLGRPWAPLTGPSVPPPSS -> MA
FT                                PRCPWPWPRRRLLDVLAPLVLLLGVRAASAEP (in
FT                                isoform Short).
FT                                /FTId=VSP_001156.
FT   VARIANT    1437   1437       D -> N (increased risk of developing
FT                                prostate cancer).
FT                                /FTId=VAR_012709.
FT   CONFLICT    428    428       F -> S (IN REF. 2).
FT   CONFLICT    841    841       I -> V (IN REF. 2).
FT   CONFLICT    877    877       V -> L (IN REF. 3).
FT   CONFLICT    886    886       P -> R (IN REF. 3).
FT   CONFLICT    912    912       P -> R (IN REF. 3).
FT   CONFLICT    933    933       R -> L (IN REF. 3).
FT   CONFLICT    975    975       P -> L (IN REF. 3).
FT   CONFLICT   1064   1064       A -> P (IN REF. 3).
FT   CONFLICT   1084   1084       L -> K (IN REF. 3).
FT   CONFLICT   1120   1120       P -> A (IN REF. 3).
FT   CONFLICT   1123   1123       P -> A (IN REF. 3).
FT   CONFLICT   1126   1126       P -> PPGP (IN REF. 2).
FT   CONFLICT   1206   1206       G -> GQ (IN REF. 3).
FT   CONFLICT   1304   1304       R -> G (IN REF. 3).
FT   CONFLICT   1314   1314       A -> G (IN REF. 3).
FT   CONFLICT   1323   1324       LR -> CG (IN REF. 3).
FT   CONFLICT   1443   1443       R -> T (IN REF. 4).
FT   CONFLICT   1483   1483       S -> Y (IN REF. 4).
SQ   SEQUENCE   1516 AA;  153840 MW;  3C70F29A4476EB76 CRC64;

Query Match             82.2%;  Score 1019;  DB 1;  Length 1516;
Best Local Similarity   82.4%;  Pred. No. 3.8e-82;
Matches  187;  Conservative  16;  Mismatches  20;  Indels  4;  Gaps  1;
```

```
Qy    1 PWRADDILAGPPRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLVA 60
         |||||||  ||| :||||||| ||||| :||||| :||||| | :|:|| ||||||||
Db 1292 PWRADDILASPPRLPEPQPYPGAPHHSSYVHLRPARPTSPP---AHSHRDFQPVLHLVA 1347

Qy   61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTPRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
         |||| ||:|||||||||||||||  | :|||||||||||||||||||||||: |:||
Db 1348 LNSPLSGGMRGIRGADFQCFQQARAVGLAGTPRAFLSSRLQDLYSIVRRADRAAVPIVNL 1407

Qy  121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180
         : |:|| ||||||||||| |||||||||||  :||||| ||||:||||||:||:|||||
Db 1408 KDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSDPNGRRLTESY 1467

Qy  181 CETWRTEAPAATGQASSLLAGRLLEQRAASCRHAFVVLCIENSVMTS 227
         ||||||||| || ||:||| |||||||  ||||| :| |||||| |:
Db 1468 CETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 1514


RESULT 2
CAIH_MOUSE
ID   CAIH_MOUSE     STANDARD;      PRT;   1774 AA.
AC   P39061; Q60672; Q61437; Q62001; Q62002; Q9JK63;
DT   01-FEB-1995 (Rel. 31, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin].
GN   COL18A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RC   STRAIN=BALB/c; TISSUE=Liver;
RX   MEDLINE=94245707; PubMed=8188673;
RA   Rehn M.V., Hintikka E., Pihlajaniemi T.;
RT   "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
RT   partial structure of the corresponding gene, and comparison of the
RT   alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
RT   chain.";
RL   J. Biol. Chem. 269:13929-13935(1994).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1, 2 AND 3).
RX   MEDLINE=96435922; PubMed=8838808;
RA   Rehn M., Hintikka E., Pihlajaniemi T.;
RT   "Characterization of the mouse collagen gene for the alpha-1 chain of type
RT   XVIII collagen (COL18A1) reveals that the three variant N-terminal
RT   polypeptide forms are transcribed from two widely separated
RT   promoters.";
RL   Genomics 32:436-446(1996).
RN   [3]
RP   SEQUENCE OF 1-1387 FROM N.A. (ISOFORM 3).
RX   MEDLINE=94240112; PubMed=8183894;
RA   Rehn M.V., Pihlajaniemi T.;
RT   "Alpha 1(XVIII), a collagen chain with frequent interruptions in the
RT   collagenous sequence, a distinct tissue distribution, and homology
RT   with type XV collagen.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RN   [4]
RP   SEQUENCE OF 487-1774 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=94240111; PubMed=8183893;
RA   Oh S.P., Kamagata Y., Muragaki Y., Timmons S., Ooshima A., Olsen B.R.;
RT   "Isolation and sequencing of cDNAs for proteins with multiple domains
RT   of Gly-Xaa-Yaa repeats identify a distinct family of collagenous
RT   proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).
RN   [5]
RP   SEQUENCE OF 1591-1774 FROM N.A.
RX   MEDLINE=21217748; PubMed=11321448;
RA   Jia S., Zhu F., Li H., He F., Xiu R.-J.;
RT   "Anticancer treatment of endostatin gene therapy by targeting tumor
RT   neovasculature in C57/BL mice.";
RL   Clin. Hemorheol. Microcirc. 23:251-257(2000).
RN   [6]
RP   CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.
RX   MEDLINE=97160848; PubMed=9008168;
RA   O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vasios G., Lane W.S.,
RA   Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;
RT   "Endostatin: an endogenous inhibitor of angiogenesis and tumor
RT   growth.";
RL   Cell 88:277-285(1997).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.
RX   MEDLINE=98169382; PubMed=9501087;
RA   Hohenester E., Sasaki T., Olsen B.R., Timpl R.;
RT   "Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A
RT   resolution.";
RL   EMBO J. 17:1656-1664(1998).
CC   -!- FUNCTION: Endostatin potently inhibits endothelial cell
CC       proliferation and angiogenesis. May inhibit angiogenesis by
CC       binding to the heparan sulfate proteoglycans involved in growth
CC       factor signaling.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative promoter;
CC         Comment=2 isoforms, 1 (shown here) and 3, are produced by use
CC         of alternative promoters;
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=1; Synonyms=NCI-764;
CC         IsoId=P39061-3; Sequence=Displayed;
CC       Name=2; Synonyms=Long, NCI-517;
CC         IsoId=P39061-1; Sequence=VSP_008303;
CC         Note=Produced by alternative splicing of isoform 1;
CC       Name=3; Synonyms=Short, NCI-301;
CC         IsoId=P39061-2; Sequence=VSP_001157, VSP_001158;
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC       unit (G-X-Y) are hydroxylated in some or all of the chains.
CC   -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC       INTERRUPTED HELICES (FACIT) FAMILY.
CC   -!- SIMILARITY: Contains 1 frizzled (FZ) domain.
CC   -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L16898; AAA37434.1; -.
DR   EMBL; U03714; AAA20657.1; -.
DR   EMBL; U03715; AAC52901.1; -.
DR   EMBL; U34606; AAC52901.1; JOINED.
DR   EMBL; U34608; AAC52901.1; JOINED.
DR   EMBL; U34609; AAC52901.1; JOINED.
DR   EMBL; U34610; AAC52901.1; JOINED.
DR   EMBL; U34611; AAC52901.1; JOINED.
DR   EMBL; U34612; AAC52901.1; JOINED.
DR   EMBL; U34613; AAC52901.1; JOINED.
DR   EMBL; U03716; AAC52901.1; JOINED.
DR   EMBL; U03718; AAC52901.1; JOINED.
DR   EMBL; U03715; AAC52902.1; -.
DR   EMBL; U34607; AAC52902.1; JOINED.
DR   EMBL; U34608; AAC52902.1; JOINED.
DR   EMBL; U34609; AAC52902.1; JOINED.
DR   EMBL; U34610; AAC52902.1; JOINED.
DR   EMBL; U34611; AAC52902.1; JOINED.
DR   EMBL; U34612; AAC52902.1; JOINED.
DR   EMBL; U34613; AAC52902.1; JOINED.
DR   EMBL; U03716; AAC52902.1; JOINED.
DR   EMBL; U03718; AAC52902.1; JOINED.
DR   EMBL; U03715; AAC52903.1; -.
DR   EMBL; U03716; AAC52903.1; JOINED.
DR   EMBL; U03718; AAC52903.1; JOINED.
DR   EMBL; U34607; AAC52903.1; JOINED.
DR   EMBL; U34608; AAC52903.1; JOINED.
```

```
DR   EMBL; U34609; AAC52903.1; JOINED.
DR   EMBL; U34610; AAC52903.1; JOINED.
DR   EMBL; U34611; AAC52903.1; JOINED.
DR   EMBL; U34612; AAC52903.1; JOINED.
DR   EMBL; U34613; AAC52903.1; JOINED.
DR   EMBL; U11636; AAC52178.1; -.
DR   EMBL; U11637; AAC52179.1; -.
DR   EMBL; L22545; AAA19781.1; -.
DR   EMBL; AF257775; AAF69009.1; -.
DR   PIR; A56101; A56101.
DR   PDB; 1KOE; 16-FEB-99.
DR   PDB; 1DY0; 11-APR-00.
DR   PDB; 1DY1; 21-JAN-01.
DR   MGD; MGI:88451; Col18a1.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   GO; GO:0001525; P:angiogenesis; IMP.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR0008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 8.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
DR   PROSITE; PS50038; FZ; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
KW   Alternative promoter usage; 3D-structure.
FT   SIGNAL        1      26       POTENTIAL.
FT   CHAIN        27    1774       COLLAGEN ALPHA 1 (XVIII) CHAIN.
FT   CHAIN      1591    1774       ENDOSTATIN.
FT   DOMAIN      245     433       TSP N-TERMINAL.
FT   DOMAIN      365     482       FZ.
FT   DOMAIN       27     785       NONHELICAL REGION 1 (NC1).
FT   DOMAIN      786     812       TRIPLE-HELICAL REGION 1 (COL1).
FT   DOMAIN      813     822       NONHELICAL REGION 2 (NC2).
FT   DOMAIN      823     896       TRIPLE-HELICAL REGION 2 (COL2).
FT   DOMAIN      897     920       NONHELICAL REGION 3 (NC3).
FT   DOMAIN      921    1042       TRIPLE-HELICAL REGION 3 (COL3).
FT   DOMAIN     1043    1065       NONHELICAL REGION 4 (NC4).
FT   DOMAIN     1066    1148       TRIPLE-HELICAL REGION 4 (COL4).
FT   DOMAIN     1149    1162       NONHELICAL REGION 5 (NC5).
FT   DOMAIN     1163    1204       TRIPLE-HELICAL REGION 5 (COL5).
FT   DOMAIN     1205    1217       NONHELICAL REGION 6 (NC6).
FT   DOMAIN     1218    1290       TRIPLE-HELICAL REGION 6 (COL6).
FT   DOMAIN     1291    1300       NONHELICAL REGION 7 (NC7).
FT   DOMAIN     1301    1333       TRIPLE-HELICAL REGION 7 (COL7).
FT   DOMAIN     1334    1345       NONHELICAL REGION 8 (NC8).
FT   DOMAIN     1346    1369       TRIPLE-HELICAL REGION 8 (COL8).
FT   DOMAIN     1370    1376       NONHELICAL REGION 9 (NC9).
FT   DOMAIN     1377    1428       TRIPLE-HELICAL REGION 9 (COL9).
FT   DOMAIN     1429    1441       NONHELICAL REGION 10 (NC10).
FT   DOMAIN     1442    1459       TRIPLE-HELICAL REGION 10 (COL10).
FT   DOMAIN     1460    1774       NONHELICAL REGION 11 (NC11).
FT   CARBOHYD    354     354       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    361     361       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    585     585       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    947     947       N-LINKED (GLCNAC...) (POTENTIAL).
FT   DISULFID   1623    1763
FT   DISULFID   1725    1755
FT   SITE       1351    1353       CELL ATTACHMENT SITE (POTENTIAL).
FT   VARSPLIC      1       1       Missing (in isoform 3).
FT                                 /FTId=VSP_001157.
FT   VARSPLIC    460     486       AGDRLPVVCASLPSQEDGYCVPIGPAA -> MAPRWHLLDV
FT                                 LTSLVLLVARVSWAEP (in isoform 3).
FT                                 /FTId=VSP_001158.
FT   VARSPLIC    240     486       Missing (In isoform 2).
FT                                 /FTId=VSP_008303.
FT   CONFLICT   1147    1147       P -> L (IN REF. 4).
FT   CONFLICT   1194    1194       P -> F (IN REF. 4).
FT   CONFLICT   1211    1211       A -> R (IN REF. 4).
```

DR   EMBL; L25286; AAA58429.1; -.
DR   EMBL; D21230; BAA04762.1; -.
DR   EMBL; L01697; -; NOT_ANNOTATED_CDS.
DR   PIR; A53317; A53317.
DR   HSSP; P39061; 1KOE.
DR   Genew; HGNC:2192; COL15A1.
DR   MIM; 120325; -.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 5.
DR   Pfam; PF02210; TSPN; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cell adhesion; Collagen; Glycoprotein; Signal.
FT   SIGNAL   1   25
FT   CHAIN   26   1388   COLLAGEN ALPHA 1(XV) CHAIN.
FT   DOMAIN   40   228   POTENTIAL.
FT   DOMAIN   229   555   TSP N-TERMINAL.
FT   DOMAIN   556   573   NONHELICAL REGION 1 (NC1).
FT   DOMAIN   574   618   TRIPLE-HELICAL REGION 1 (COL1).
FT   DOMAIN   619   732   NONHELICAL REGION 2 (NC2).
FT   DOMAIN   733   763   TRIPLE-HELICAL REGION 2 (COL2).
FT   DOMAIN   764   798   NONHELICAL REGION 3 (NC3).
FT   DOMAIN   799   822   TRIPLE-HELICAL REGION 3 (COL3).
FT   DOMAIN   823   867   NONHELICAL REGION 4 (NC4).
FT   DOMAIN   868   878   TRIPLE-HELICAL REGION 4 (COL4).
FT   DOMAIN   879   949   NONHELICAL REGION 5 (NC5).
FT   DOMAIN   950   983   TRIPLE-HELICAL REGION 5 (COL5).
FT   DOMAIN   984   1013   NONHELICAL REGION 6 (NC6).
FT   DOMAIN   1014   1027   TRIPLE-HELICAL REGION 6 (COL6).
FT   DOMAIN   1028   1045   NONHELICAL REGION 7 (NC7).
FT   DOMAIN   1046   1052   TRIPLE-HELICAL REGION 7 (COL7).
FT   DOMAIN   1053   1107   NONHELICAL REGION 8 (NC8).
FT   DOMAIN   1108   1117   TRIPLE-HELICAL REGION 8 (COL8).
FT   DOMAIN   1118   1132   NONHELICAL REGION 9 (NC9).
FT   DOMAIN   1133   1388   TRIPLE-HELICAL REGION 9 (COL9).
FT   DOMAIN   358   555   NONHELICAL REGION 10 (NC10).
FT   REPEAT   358   408   4 X TANDEM REPEATS.
FT   REPEAT   409   459   1.
FT   REPEAT   460   509   2.
FT   REPEAT   510   555   3.
FT   CARBOHYD   306   306   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   324   324   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   687   687   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   807   807   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   814   814   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1046   1046   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT   10   10   C -> S (IN REF. 2).
FT   CONFLICT   49   49   D -> V (IN REF. 2).
FT   CONFLICT   95   95   L -> A (IN REF. 2).
FT   CONFLICT   150   150   P -> A (IN REF. 2).
FT   CONFLICT   204   204   M -> V (IN REF. 2).
FT   CONFLICT   409   409   R -> A (IN REF. 2).
SQ   SEQUENCE   1388 AA; 141930 MW; 60822AD925A3093D CRC64;

Query Match   42.7%; Score 528.5; DB 1; Length 1388;
Best Local Similarity   48.2%; Pred. No. 8.5e-39;
Matches 106; Conservative 30; Mismatches 63; Indels 21; Gaps   3;

Qy   11 PPRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHQDPQLVLHLVALNSPQPGGMR 70
                 |||||||| |:||| : ::||: ||| ||| :: |: || |||| |: | ||
Db   1190 PPPALSSNFHQLLP--------PPNPISSANYE-------KPALHLAALNMPFSGDIR 1232

Qy   71 GIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNLRDEVLFPSWE 130
                 |||:||||||| | |:||| |||| ||||||:| :|::|| : :|||| :| :|: |:

Qy   191 ATGQASSLLAGRLLEQBAASCRHAPVVLCIENSVMTSFSK 230
             | : :|||||| :: |:| |:|||:|||||:|::| |
Db   1349 VTGLASPLSTGKILDQKAYSCANRLIVLCIENSFMTDARK 1388

RESULT 4
THIC_MYCLE
ID   THIC_MYCLE   STANDARD;   PRT;   547 AA.
AC   Q9ZBL0;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Thiamine biosynthesis protein thiC.
GN   THIC OR ML0294 OR MLCB1450.28C.
OS   Mycobacterium leprae.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M. Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
CC   -!- FUNCTION: Required for the synthesis of the hydromethylpyrimidine
CC   (HMP) moiety of thiamine (4-amino-2-methyl-5-
CC   hydroxymethylpyrimidine) (By similarity).
CC   -!- PATHWAY: Thiamine biosynthesis.
CC   -!- SIMILARITY: Belongs to the thiC family.
CC   ---
DR   EMBL; AL035159; CAA22712.1; -.
DR   EMBL; AL583918; CAC29802.1; -.
DR   PIR; T44743; T44743.
DR   Leproma; ML0294; -.
DR   HAMAP; MF_00089; -; 1.
DR   InterPro; IPR002817; ThiC.
DR   Pfam; PF01964; ThiC; 1.
DR   ProDom; PD007048; ThiC; 1.
DR   TIGRFAMs; TIGR00190; thiC; 1.
KW   Thiamine biosynthesis; Complete proteome.
SQ   SEQUENCE   547 AA; 59840 MW; 797DFB09DE462D58 CRC64;

Query Match   7.5%; Score 93.5; DB 1; Length 547;
Best Local Similarity 27.4%; Pred. No. 0.87;
Matches 43; Conservative 14; Mismatches 49; Indels 51; Gaps   8;

Qy   19 PYPGAPHHGSYVHFQPARPTGGP--------VHTHTHQDF------------QLVLHLVA 60
              :| ||||::|: | |: | | : | : || ::|| |:||
Db   17 PIFGS---SKAYREVANPDGGPSLRVPFRRVHLSTGAHFDLYDTSGPVTDPDAVINLTA 72

```
Qy   61 LNSPQPGGWRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
            |||·|·|::||||·||·||···||··||···||:·||·||
Db   73 GLPPRPGVIRD-RGTQ---LQRARAGEITAEM-AFI-----------ADREGMPAELV 114

Qy  121 RDEVLF------------PSWEALFSGSEGQLKPGARI 146
                              :·|·|
Db  115 RVEVALGRAVIPANHNHPBIEPMIIGKAFAVKVNANI 151


RESULT 5
GLI_CHICK
ID   GLI_CHICK     STANDARD;       PRT;    556 AA.
AC   P55878;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein GLI1 (GLI) (Fragment).
GN   GLI1 OR GLI.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97105842; PubMed=8948590;
RA   Marigo V., Johnson R.L., Vortkamp A., Tabin C.J.;
RT   "Sonic hedgehog differentially regulates expression of GLI and GLI3
RT   during limb development.";
RL   Dev. Biol. 180:273-283(1996).
CC   -!- FUNCTION: MAY REGULATE THE TRANSCRIPTION OF SPECIFIC GENES DURING
CC       NORMAL DEVELOPMENT. MAY PLAY A ROLE IN CRANIOFACIAL DEVELOPMENT
CC       AND DIGITAL DEVELOPMENT, AS WELL AS DEVELOPMENT OF THE CENTRAL
CC       NERVOUS SYSTEM AND GASTROINTESTINAL TRACT. IMPLICATED IN THE
CC       TRANSDUCTION OF SHH SIGNAL (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   --------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------
DR   EMBL; U60762; AAB51659.1; -.
DR   HSSP; P08151; 2GLI.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 5.
DR   SMART; SM00355; ZnF_C2H2; 5.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 5.
KW   Zinc-finger; Metal-binding; DNA-binding; Transcription regulation;
KW   Nuclear protein; Repeat.
FT   ZN_FING    247    272    C2H2-TYPE.
FT   ZN_FING    280    307    C2H2-TYPE.
FT   ZN_FING    313    337    C2H3-TYPE.
FT   ZN_FING    343    368    C2H2-TYPE.
FT   ZN_FING    374    399    C2H2-TYPE.
FT   NON_TER    556    556
SQ   SEQUENCE  556 AA;  60215 MW;  722D2DA5A1CA4D98 CRC64;

Query Match  7.3%;  Score 90;  DB 1;  Length 556;
Best Local Similarity  26.0%;  Pred. No. 1.8;
Matches  63;  Conservative  22;  Mismatches  93;  Indels  64;  Gaps  13;

Qy   17 PQPYPGAPH------HGSVVHFQPARPTGGPVHTH---THTHQDFQLVLHLVALNSPQP 66
           ||·|                                 :·||
Db    7 PQARPYAEHCPRPIHGASAGTPGLQGLDFPV-CHQPNLASSHHGYGLV-----PGTBHP 60

Qy   67 GGMRGIRGADFQCFQQARAAGLAGTFRAF----LSSRLQDLYSIVRRADRTGVPVVNLRD 122
```

```
Db   61 GG------AADGSRFSTPRGAGXLGKKRALSISPLSDSSVDLQTVIRTSPNSLVAFINSRC 115
Qy  123 EVLFPSWEAL-----------PSGSEGQLKPGARIPS------------FDGRDVLQHPA 159
           :·|·|                       |·|
Db  116 ASAGGSYGHLSISTISPSLGYQNPFGQQKGQGQLRSHTPLPPCSSHETLSSRPGLLHPT 175
Qy  160 WPRKSVWHGSDPSGRR-----LTDSYCETWRTE----APAATQOASSLLAGRL-----LE 205
           ||                    :·||·|               :·||:|::|
Db  176 PARGTIKHCQQLKLERSLSSPLTAKYPEB-KSEGDISSPASTGTQDPLL-GMLSVRDDLE 233
Qy  206 QE 207
           :|
Db  234 KE 235


RESULT 6
NME3_HUMAN
ID   NME3_HUMAN     STANDARD;        PRT;    1233 AA.
AC   Q14957;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Glutamate [NMDA] receptor subunit epsilon 3 precursor (N-methyl
DE   D-aspartate receptor subtype 2C) (NR2C) (NMDAR2C).
GN   GRIN2C.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97189248; PubMed=9037519;
RA   Lin Y.J., Bovetto S., Carver J.M., Giordano T.;
RT   "Cloning of the cDNA for the human NMDA receptor NR2C subunit and its
RT   expression in the central nervous system and periphery.";
RL   Brain Res. Mol. Brain Res. 43:57-64(1996).
CC   -!- FUNCTION: NMDA receptor subtype of glutamate-gated ion channels
CC       with high calcium permeability and voltage-dependent sensitivity
CC       to magnesium. Mediated by glycine.
CC   -!- SUBUNIT: Heterodimer of an epsilon subunit and a zeta subunit.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Mainly in brain with predominant expression is
CC       in the cerebellum, also present in the hippocampus, amygdala,
CC       caudate nucleus, corpus callosum, subthalanic nuclei and thalamus.
CC       Detected in the heart, skeletal muscle and pancreas.
CC   -!- SIMILARITY: Belongs to the ligand-gated ionic channel family.
CC   --------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------
DR   EMBL; L76224; AAA88096.1; -.
DR   HSSP; P19491; 1GR2.
DR   Genew; HGNC:4587; GRIN2C.
DR   MIM; 138254; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0004972; F:N-methyl-D-aspartate selective glutamate re. . .; TAS.
DR   GO; GO:0007215; P:glutamate signaling pathway; TAS.
DR   GO; GO:0006810; P:transport; TAS.
DR   InterPro; IPR001828; ANF_receptor.
DR   InterPro; IPR001320; Ion_glu_receptor.
DR   InterPro; IPR001508; NMDA_receptor.
DR   InterPro; IPR001111; SBP/Glu_receptor.
DR   Pfam; PF01094; ANF_receptor; 1.
DR   Pfam; PF00060; lig_chan; 1.
DR   PRINTS; PR00177; NMDARECEPTOR.
DR   SMART; SM00079; PBPe; 1.
```

```
KW   Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;
KW   Ionic channel; Magnesium.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1233       GLUTAMATE [NMDA] RECEPTOR SUBUNIT
FT                                EPSILON 3.
FT   TRANSMEM    554    574       1 (POTENTIAL).
FT   TRANSMEM    597    617       2 (POTENTIAL).
FT   TRANSMEM    627    647       3 (POTENTIAL).
FT   TRANSMEM    815    835       4 (POTENTIAL).
FT   SITE        612    612       FUNCTIONAL DETERMINANT OF NMDA
FT                                RECEPTORS (BY SIMILARITY).
FT   CARBOHYD     70     70       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    337    337       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    438    438       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    539    539       N-LINKED (GLCNAC...)  (POTENTIAL).
SQ   SEQUENCE   1233 AA;  134239 MW;  700F24629ED0D699 CRC64;

Query Match         7.1%;  Score 88;  DB 1;  Length 1233;
Best Local Similarity  26.6%;  Pred. No. 6.8;  Indels  66;  Gaps  12;
Matches   63;  Conservative  15;  Mismatches  93;

Qy      9  AGPPRLL--DPQPYPGAPHHGSVHFQPARPTGGPVHTHTHQDPQLVLHIVALNSPQ   65
              ||:||                                        :
Db    940  SGPSPCLPTPDPPPEPSPTGWGP-----------------------ALVRRAPQ  976

Qy     66  PGGMRGIRGADFQCFQ-------QARAAGLAGTFRAFLSSRLQDL-----YSIVRRAD  111
              ||                    |||                  ||
Db    977  PPGRPPTPGPPLSDVSRVSRRPAWEARWPVRTGHCGRHLSASERPLSPARCHYSSFPRAD 1036

Qy    112  RTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDP 171
              |:::|||                   ||
Db   1037  RSGRPFL-----PLPELEDLPLLGKEQL---AR-------REALLHAAWAR------GSRP 1077

Qy    172  SGRRLTDSYCETWR--TEAPA-ATGQASSLLAG---RLLEQEAASCRHAFVVLCIE  221
              |:  | |          |    |::||||
Db   1078  RHASLPSSVAEAFARPSSLPAGCTGPACARPDGHSACRRLAQAQSMCLPIYREACQE 1134

RESULT 7
ATX1_HUMAN
ID   ATX1_HUMAN     STANDARD;      PRT;    816 AA.
AC   P54253; Q9UJG2; Q9Y4J1;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Ataxin-1 (Spinocerebellar ataxia type 1 protein).
GN   SCA1 OR ATX1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Cerebellum;
RX   MEDLINE=95038838; PubMed=7951322;
RA   Banfi S., Servadio A., Chung M.-Y., Kwiatkowski T.J. Jr., McCall A.E.,
RA   Duvick L.A., Shen Y., Roth E.J., Orr H.T., Zoghbi H.Y.;
RT   "Identification and characterization of the gene causing type 1
RT   spinocerebellar ataxia";
RL   Nat. Genet. 7:513-519(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Tubby B.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 189-231 FROM N.A.
RX   MEDLINE=96177682; PubMed=8634720;
RA   Quan F., Janas J., Popovich B.W.;
RT   "A novel CAG repeat configuration in the SCA1 gene: implications for
RT   the molecular diagnostics of spinocerebellar ataxia type 1.";
RL   Hum. Mol. Genet. 4:2411-2413(1995).
RN   [4]
RP   SELF-ASSOCIATION SITE.
RX   MEDLINE=97252384; PubMed=9097953;
RA   Burright E.N., Davidson J.D., Duvick L.A., Koshy B., Zoghbi H.Y.,
RA   Orr H.T.;
RT   "Identification of a self-association region within the SCA1 gene
RT   product, ataxin-1.";
RL   Hum. Mol. Genet. 6:513-518(1997).
RN   [5]
RP   RNA-BINDING DOMAIN.
RX   MEDLINE=21065637; PubMed=11136710;
RA   Yue S., Serra H.G., Zoghbi H.Y., Orr H.T.;
RT   "The spinocerebellar ataxia type 1 protein, ataxin-1, has RNA-binding
RT   activity that is inversely affected by the length of its polyglutamine
RT   tract.";
RL   Hum. Mol. Genet. 10:25-30(2001).
CC   -!- FUNCTION: Binds RNA in vitro. May be involved in RNA metabolism.
CC       The expansion of the polyglutamine tract may alter this function.
CC   -!- SUBUNIT: Interacts with LANP and A1U.
CC   -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=1;
CC         Comment=At least 2 isoforms are produced;
CC       Name=1;
CC         IsoId=P54253-1; Sequence=Displayed;
CC   -!- TISSUE SPECIFICITY: Widely expressed throughout the body.
CC   -!- POLYMORPHISM: The poly-Gln region of SCA1 is highly polymorphic (4
CC       to 39 repeats) in the normal population and is expanded to about
CC       40-83 repeats in sca1 patients. Longer expansions result in
CC       earlier onset and more severe clinical manifestations of the
CC       disease.
CC   -!- DISEASE: Defects in SCA1 are the cause of spinocerebellar ataxia
CC       type 1 (SCA1) [MIM:164400]; also known as olivopontocerebellar
CC       atrophy I (OPCA I). SCA1 is an autosomal dominant
CC       neurodegenerative disorder characterized by progressive neuronal
CC       loss in the cerebellum, brain stem and spinocerebellar tracts.
CC       Clinical features are cerebellar ataxia, dysarthria,
CC       ophtalmoparesis, muscle wasting and neuropathy. Onset of the
CC       disease usually occurs in the third or fourth decade of life and
CC       death occurs ten to twenty years later.
CC   -!- MISCELLANEOUS: The self-association seems to be necessary to form
CC       nuclear aggregates.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X79204; CAA57793.1; -.
DR   EMBL; AL009031; CAA15622.1; -.
DR   EMBL; S82497; AAD14401.1; -.
DR   PIR; S46268; S46268.
DR   Genew; HGNC:10548; SCA1.
DR   MIM; 601556; -.
DR   MIM; 164400; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   InterPro; IPR003652; Ataxin.
DR   SMART; SM00536; AXH; 1.
KW   RNA-binding; Nuclear protein; Polymorphism; Triplet repeat expansion;
KW   Alternative splicing.
FT   DOMAIN       197    226       POLY-GLN.
FT   DOMAIN       495    605       SELF-ASSOCIATION SITE.
FT   DOMAIN       541    767       RNA BINDING.
FT   DOMAIN       795    798       NUCLEAR LOCALIZATION SIGNAL
FT                                 (BY SIMILARITY).
SQ   SEQUENCE    816 AA;   87051 MW;  D49BA5DB423D0777 CRC64;

Query Match         7.1%;  Score 87.5;  DB 1;  Length 816;
Best Local Similarity  22.4%;  Pred. No. 4.7;
Matches   54;  Conservative  26;  Mismatches  96;  Indels  65;  Gaps  10;
```

```
Qy     8 LAGPPRLLDPQPYPGAPHHGSYVHP--------QPARPTGGPVHTHTHQDFQLVLHLV  59
               ||:           |:|:           :|:        ::     ::|
Db   228 LSRAPGLITPGSPPPA-QQNQYVHISSSPQNTGRTASPPAIPVHLHPHQ---TMIPHTL 282

Qy    60 ALNSPPPGGMR-GIRGADPQCFQQARAAGLAGTFRAFLSSRLQD-------LYSIVRRAD 111
             :|:|        |:|:                                  ||||
Db   283 TLGPPSQVVMQYADSGSHFVPREATKKAE--------SSRLQQAIQAKEVLNGEMEKSR 333

Qy   112 RTGVPVVNLRDEVLFPSWEALFPSSEGQLKPGARI-FSFDGRDVLQHPAWPRKSVWHGS 169
            |               :|::|  :|:|||  :|||      |:|:    | :
Db   334 RYGAP-----------SSADLGLGKAGGKSVPHYESRHVVVHFS---PSDYSSR 374

Qy   170 DPSGRRLT--------DSYCETWRTEAPAATGQASLLAGRLLEQEAASCRHAF 215
           ||  ||:         :||  |:  :|                  |:|
Db   375 DPSGVRASVVNVLPNSNTPAADLEVQQATHREASPSTLNDKSGLHLGKPGHRSYALSPHTV 434

Qy   216 V 216
           :
Db   435 I 435


RESULT 8
ALG8_PSEAE
ID   ALG8_PSEAE     STANDARD;      PRT;     494 AA.
AC   Q52463; Q9HY70;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Glycosyl transferase alg8 (EC 2.4.1.-).
GN   ALG8 OR PA3541.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=8830;
RX   MEDLINE=9124011; PubMed=8294014;
RA   Maharaj R., May T.B., Wang S.-K., Chakrabarty A.M.;
RT   "Sequence of the alg8 and alg44 genes involved in the synthesis of
RT   alginate by Pseudomonas aeruginosa.";
RL   Gene 136:267-269(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saler M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- FUNCTION: Possibly a processive enzyme that polymerizes GDP-
CC       mannuronic acid.
CC   -!- PATHWAY: Alginate biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 2.
```

```
DR   EMBL; L22611; AAC35875.2; ALT_INIT.
DR   EMBL; AE004774; AAG06929.1; -.
DR   PIR; A83204; A83204.
KW   Alginate biosynthesis; Transferase; Glycosyltransferase;
```

```
KW   Transmembrane; Complete proteome.
FT   TRANSMEM      12     34       POTENTIAL.
FT   TRANSMEM      49     71       POTENTIAL.
FT   TRANSMEM     381    403       POTENTIAL.
FT   TRANSMEM     423    445       POTENTIAL.
FT   CONFLICT     115    115       A -> R (IN REF. 1).
SQ   SEQUENCE   494 AA;   56456 MW;  D89A5627E913FDF5 CRC64;

Query Match              7.0%;  Score 87;  DB 1;  Length 494;
Best Local Similarity   24.6%;  Pred. No. 2.9;
Matches   45;  Conservative  20; Mismatches  68; Indels  50; Gaps   7;

Qy    50 QDFQLVLHLVALNSPQPGGMRGIRGADF--------QCFQQARAAGLAGTFRAFLSSR  99
           ||:|| :|:|  |: |:|:  |  |:            :|:|      |:|  |::
Db    39 KDFILLIGAVGIWRYSMGGVHFLRGMLFLHVVYPYYRRVRQLGSAADPSHVFLMVTSFR  98

Qy   100 LQDL------YSIVRRADRTGVP--------VVNLRDEVLFPS-WEAL------------ 132
                     |:|  : :|               : |   : |
Db    99 IDALTTAMVRSVIREAIDSGYPTTVVCSIVEMSDEVLVRSLWEKMNPPDRVSLDFVRIP 158

Qy   133 FSGSEQLKPGARIFS--------------FDGRDVLQHPAWPRKSVWHGSDPS-GRRLTDS 179
             :| :  |                     ||:|   |::    :    |:  |:
Db   159 GTGKRDGLAYGFRAISRHLPDDDAVVAVIDGDTVLDHGVVKKTVPWFKLPPNVGGLITNE 218

Qy   180 YCE 182
           :||
Db   219 FCE 221


RESULT 9
HGD_BRAJA
ID   HGD_BRAJA      STANDARD;      PRT;     448 AA.
AC   Q89XH1;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Homogentisate 1,2-dioxygenase (EC 1.13.11.5) (Homogentisicase)
DE   (Homogentisate oxygenase) (Homogentisic acid oxidase).
GN   HMGA OR BLL0343.
OS   Bradyrhizobium japonicum.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bradyrhizobiaceae; Bradyrhizobium.
OX   NCBI_TaxID=375;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=USDA 110;
RX   MEDLINE=22484998; PubMed=12597275;
RA   Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA   Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA   Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA   Tabata S.;
RT   "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT   Bradyrhizobium japonicum USDA110.";
RL   DNA Res. 9:189-197(2002).
CC   -!- CATALYTIC ACTIVITY: Homogentisate + O(2) = 4-maleylacetoacetate.
CC   -!- COFACTOR: Iron (By similarity).
CC   -!- PATHWAY: Catabolism of tyrosine; third step.
CC   -!- PATHWAY: Catabolism of phenylalanine; fourth step.
CC   -!- SIMILARITY: Belongs to the homogentisate dioxygenase family.
```

```
DR   EMBL; AP005936; BAC45608.1; ALT_INIT.
DR   HAMAP; MF_00334; -; 1.
DR   InterPro; IPR007113; Cupin_sup.
DR   InterPro; IPR005708; HmgA.
DR   Pfam; PF04209; HgmA; 1.
```

DR   TIGRFAMS; TIGR01015; hmgA; 1.
KW   Oxidoreductase; Dioxygenase; Metal-binding; Iron;
KW   Phenylalanine catabolism; Tyrosine catabolism; Complete proteome.
FT   METAL   346   346   IRON (BY SIMILARITY).
FT   METAL   352   352   IRON (BY SIMILARITY).
FT   METAL   382   382   IRON (BY SIMILARITY).
SQ   SEQUENCE   448 AA; 49780 MW; F6DD46SE68735D3C CRC64;

Query Match   6.9%; Score 86; DB 1; Length 448;
Best Local Similarity 24.0%; Pred. No. 3.2;
Matches 44; Conservative 25; Mismatches 78; Indels 36; Gaps 9;

Qy   24 PH-------HGSVHFQPARPTGGPVHTHTHTHQDFQLVLHLVALNSPQPGGMRGIRGA 75
Db   270 PHSPIDVVAWHGNYAPYKYDLRTFSPVGAIGFDHPDPSIFTVLTS----PSETAGTANI 324

Qy   76 DFQCFQQARAAGLAGTFRA------FLSSRLQDLYSIVRRADRTGVP-VVNLRDEVL--F 126
Db   325 DFVIFPE-RMWVADNTFRPPWYHMNIMSEFMGLHIYGVVDAKPQGFVPGGMSLHNCMLPHG 383

Qy   127 PSWEALFSGSEGQLKP----GARIFSFPDGRDRVLQHPAMPRKSVWHGHWGSDPSGRRLITDSYCE 182
Db   384 PDRDAFEHASNGELKPVKLPVKLLTGTMAFMFEIR------YPQRVTAHAANAS--TLQDDYAD 434

Qy   183 TWR 185
Db   435 CWK 437

RESULT 10
NRT1_CHICK
ID   NRT1_CHICK   STANDARD;   PRT;   312 AA.
AC   P55806;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   NAD(P)(+)--arginine ADP-ribosyltransferase 1 precursor (EC 2.4.2.31)
DE   (Mono(ADP-ribosyl)transferase 1) (AT1).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=White leghorn; TISSUE=Bone marrow;
RX   MEDLINE=95050487; PubMed=7961658;
RA   Tsuchiya M., Hara N., Yamada K., Osago H., Shimoyama M.;
RT   "Cloning and expression of cDNA for arginine-specific ADP-
RT   ribosyltransferase from chicken bone marrow cells.";
RL   J. Biol. Chem. 269:27451-27457(1994).
CC   -!- CATALYTIC ACTIVITY: NAD(+) + L-arginine = nicotinamide + N(2)-
CC   (ADP-D-ribosyl)-L-arginine.
CC   -!- SUBCELLULAR LOCATION: THE MATURE ENZYME IS PROBABLY SECRETED
CC   FROM GRANULOCYTES INTO THE EXTRACELLULAR SPACE.
CC   -!- SIMILARITY: Belongs to the Arg-specific ADP-ribosyltransferase
CC   family.
DR   EMBL; D31864; BAA06664.1; -.
DR   PIR; A55461; A55461.
DR   InterPro; IPR000768; ART.
DR   Pfam; PF01129; ART; 1.
DR   PRINTS; PR00970; RIBTRNSFRASE.
DR   PROSITE; PS01291; ART; 1.
KW   Transferase; Glycosyltransferase; NAD; Signal; Zymogen.

FT   SIGNAL   1   20   POTENTIAL.
FT   PROPEP   21   31   NAD(P)(+)--ARGININE ADP-
FT   CHAIN   32   266   RIBOSYLTRANSFERASE 1.
FT   POTENTIAL.
FT   PROPEP   267   312   POTENTIAL.
FT   ACT_SITE   231   231   BY SIMILARITY.
SQ   SEQUENCE   312 AA; 35318 MW; B82980439BC904FC CRC64;

Query Match   6.9%; Score 85; DB 1; Length 312;
Best Local Similarity 26.5%; Pred. No. 2.6;
Matches 53; Conservative 15; Mismatches 62; Indels 70; Gaps 13;

Qy   32 FQPARPTGGPVH-THTHTHQ----DFQLVLHLVALNSPQP------GGMRGIRGADFQCF 80
Db   118 FNEATRQGGRSHQDYIHSYHFKTLHFFLTQALFALRASQPRCYYVRGVRGIR------F 171

Qy   81 QQARAAGLAGTFRAFLSSRL-------QDLYSIVRRADRTGVPVVNL----RDEVLF 126
Db   172 MTQR-GKSVRFGQFTSTSLRKEATVNFGQDTLFVVKTC--YGVPIKQFSFPPSEDEVLI 227

Qy   127 PSWEAL-----FSGSEGQ-------------------LKPGARIFSFPDGRDV-----LQH 157
Db   228 PPFEVFEVINFSNDRGSVKIQLHSKGKMSTHNCELLKPQGGQWGRGHQEVGLGLSPGLSL 287

Qy   158 PAWP-RKSVWHG-----SDP 171
Db   288 PVLPCRRRVWEGLGHREGDP 307

RESULT 11
CHRD_MOUSE
ID   CHRD_MOUSE   STANDARD;   PRT;   948 AA.
AC   Q9ZEE2;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Chordin precursor.
GN   CHRD.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B6JL/Fl;
RA   Lu B., Bachiller D., Agius E., Piccolo S., De Robertis E.M.;
RT   "BMP-binding domains in the chordin secreted protein.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99000848; PubMed=9782094;
RA   Pappano W.N., Scott I.C., Clark T.G., Eddy R.L., Shows T.B.,
RA   Greenspan D.S.;
RT   "Coding sequence and expression patterns of mouse chordin and mapping
RT   of the cognate mouse chrd and human CHRD genes.";
RL   Genomics 52:236-239(1998).
CC   -!- FUNCTION: Dorsalizing factor. Key developmental protein that
CC   dorsalizes early vertebrate embryonic tissues by binding to
CC   ventralizing TGF-beta family bone morphogenetic proteins (BMPs)
CC   and sequestering them in latent complexes.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- DEVELOPMENTAL STAGE: Detected at high levels of a in 7 dpc mouse
CC   embryos; its level decrease at later developmental stages and in
CC   adult tissues.
CC   -!- SIMILARITY: Belongs to the chordin family.
CC   -!- SIMILARITY: Contains 4 CHRD domains.
CC   -!- SIMILARITY: Contains 4 VWFC domains.

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------
DR   EMBL; AF096276; AAD19895.1; -.
DR   EMBL; AF069501; AAC88867.1; -.
DR   MGD; MGI:1313268; Chrd.
DR   GO; GO:0007420; P:brain development; IMP.
DR   GO; GO:0007389; P:pattern specification; IMP.
DR   GO; GO:0001501; P:skeletal development; IMP.
DR   InterPro; IPR006559; SOG.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF00093; vwc; 4.
DR   SMART; SM00566; SOG; 3.
DR   SMART; SM00214; VWC; 4.
DR   PROSITE; PS50933; CHRD; 4.
DR   PROSITE; PS01208; VWFC_1; 2.
DR   PROSITE; PS50184; VWFC_2; 2.
KW   Developmental protein; Repeat; Glycoprotein; Signal.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27    948       CHORDIN.
FT   DOMAIN       49    126       VWFC 1.
FT   DOMAIN      168    277       CHRD 1.
FT   DOMAIN      279    398       CHRD 2.
FT   DOMAIN      399    520       CHRD 3.
FT   DOMAIN      526    646       CHRD 4.
FT   DOMAIN      699    759       VWFC 2.
FT   DOMAIN      779    845       VWFC 3.
FT   DOMAIN      867    927       VWFC 4.
FT   CARBOHYD    347    347       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    877    877       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   948 AA;  101512 MW;  4DC2DA01D9BD2147 CRC64;

Query Match             6.8%;  Score 84;  DB 1;  Length 948;
Best Local Similarity   24.2%;  Pred. No. 11;
Matches   53;  Conservative   17;  Mismatches   79;  Indels   70;  Gaps    9;

Qy   15 LDPQP------YPGAPHHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLVALNSPQPGG 68
            |||
Db  132 LDPQPAGLVFEYPRDPEHRSY--SDRGEPGVGERTRADGHTDF-----VAL------- 175

Qy   69 MRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNLRDEVLFPS 128
                :||    ||    :|:|          :|:|         :||:
Db  176 ---LITGPRSQAVARARVSLLRSSLRFSVSYQRLDRPSRVRFTDPTG------NILF- 222

Qy  129 WEALFSGSEGQLKFGARIFSPDGRDVLQHPAMPRK------SVWHGGSDPSGSGRRLTDSYCET 183
                :|||  |                 :|||:||
Db  223 -----------------------EHPATPTQDGLVCGVWRAVPRLSVRLLRA--EQ 253

Qy  184 WRTEAPAATCQASSLLAGRLLEQEAASCRHAFVVLCIEN 222
            :|::  |:  |:|  |  :|:|  :|:|
Db  254 LRVALVTST-HPSGEVWGPLIWQGALAAETFSAILTLED 291

RESULT 12
HTS1_COCCA
ID   HTS1_COCCA     STANDARD;      PRT;   5217 AA.
AC   Q01886;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   HC-toxin synthetase (EC 6.3.2.-) (HTS).
GN   HTS1.
OS   Cochliobolus carbonum (Bipolaris zeicola).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC   Pleosporales; Pleosporaceae; Cochliobolus.
OX   NCBI_TaxID=5017;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=ATCC 90305 / SB111;
RX   MEDLINE=93100328; PubMed=1281482;
RA   Scott-Craig J.S., Panaccione D.G., Pocard J.-A., Walton J.D.;
RT   "The cyclic peptide synthetase catalyzing HC-toxin production in the
RT   filamentous fungus Cochliobolus carbonum is encoded by a
RT   15.7-kilobase open reading frame.";
RL   J. Biol. Chem. 267:26044-26049(1992).
RN   [2]
RP   FUNCTION.
RC   STRAIN=ATCC 90305 / SB111;
RX   MEDLINE=20138231; PubMed=10671527;
RA   Cheng Y.-Q., Walton J.D.;
RT   "A eukaryotic alanine racemase gene involved in cyclic peptide
RT   biosynthesis.";
RL   J. Biol. Chem. 275:4906-4911(2000).
CC   -!- FUNCTION: Non-ribosomal peptide synthetase, able to activate
CC       proline and AEO (2-amino-9,10-epoxi-8-oxodecanoic acid), and
CC       epimerize L-Pro. Catalyzes the production of HC-toxin: a cyclic
CC       tetrapeptide. Activates and thioesterifies L-Pro, and epimerizes
CC       it to D-Pro; also uses D-Ala as a substrate but this is epimerized
CC       from L-Ala by TOXG.
CC   -!- COFACTOR: Contains 4 covalently bound phosphopantetheines.
CC   -!- PATHWAY: HC-toxin non-ribosomal biosynthesis.
CC   -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC       family.
CC   -!- SIMILARITY: Contains 4 acyl carrier domains.
CC   -!- CAUTION: It is uncertain whether Met-1 or Met-3 is the initiator.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M98024; AAA33023.1; -.
DR   HSSP; P14687; 1AMU.
DR   GO; GO:0016881; F:acid-D-amino acid ligase activity; IDA.
DR   GO; GO:0009403; P:toxin biosynthesis; IDA.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR006163; Pp_bind.
DR   InterPro; IPR006162; Ppantne_S.
DR   Pfam; PF00501; AMP-binding; 4.
DR   Pfam; PF00668; Condensation; 5.
DR   Pfam; PF00550; pp-binding; 4.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 4.
DR   PROSITE; PS00455; AMP_BINDING; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
KW   Ligase; Multifunctional enzyme; Phosphopantetheine; Repeat.
FT   REPEAT      249    842       DOMAIN 1.
FT   REPEAT     1854   2452       DOMAIN 2.
FT   REPEAT     3006   3606       DOMAIN 3.
FT   REPEAT     4158   4738       DOMAIN 4.
FT   DOMAIN      762    840       ACYL CARRIER (ACP) 1.
FT   DOMAIN     2384   2450       ACYL CARRIER (ACP) 2.
FT   DOMAIN     3536   3604       ACYL CARRIER (ACP) 3.
FT   DOMAIN     4667   4736       ACYL CARRIER (ACP) 4.
FT   BINDING     803    803       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    2414   2414       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    3568   3568       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    4700   4700       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE  5217 AA;  574637 MW;  033ID9C5400163A5 CRC64;

Query Match             6.7%;  Score 82.5;  DB 1;  Length 5217;
Best Local Similarity   24.9%;  Pred No. 1.1e+02;
Matches   43;  Conservative   15;  Mismatches   72;  Indels   43;  Gaps    8;

Qy    1 PWRADDILAGPRLLDPQPVPGAPHHGSVHFQPARPTGGPVHTHTHQDFQLVLHLVA 60
            |:| ::      |    :|:|
Db 1713 PKRPTNVKAGEPE----------HHLCPEELSMSDPT---------EPDLTLFV-- 1747

Qy   61 LNSPQPGGVMRGIRGADFQC--FQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVV 118
            |:|      ||:   :|:|    |:
Db 1748 --EPTQESNEVSAHLDFKLSYISQAYATSIASTVAHILSELVHDPY---RALNT-LPIV 1800
```

```
Qy    119 NLRDEVLFPSWEALFSGSEGQLKPGAR--IFSFDGRDVLQHPAWPRKSVWHGS 169
              |||  | |   |        | |          ||  ||
Db   1801 SEHDTAIIRSW-------NDHLFPPATECIHETFSRKVVEHPQREAICSWDGS 1846
```

RESULT 13
EGR4_HUMAN
```
ID   EGR4_HUMAN     STANDARD;      PRT;   486 AA.
AC   Q05215;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Early growth response protein 4 (EGR-4) (AT133).
GN   EGR4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9278383; PubMed=8504297;
RA   Holst C., Skerka C., Lichter P., Bialonski A., Zipfel P.F.;
RT   "Genomic organization, chromosomal localization and promoter function
RT   of the human zinc-finger gene pAT133.";
RL   Hum. Mol. Genet. 2:367-372(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92052214; PubMed=1658795;
RA   Mueller H.-J., Skerka C., Bialonski A., Zipfel P.F.;
RT   "Clone pAT133 identifies a gene that encodes another human member of
RT   a class of growth factor-induced genes with almost identical zinc-
RT   finger domains.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:10079-10083(1991).
```
```
DR   EMBL; X69438; CAA49214.1; -.
DR   EMBL; X60104; CAA42698.1; -.
DR   PIR; A41537; A41537.
DR   HSSP; P08046; 1AAY.
DR   TRANSFAC; T05190; -.
DR   Genew; HGNC:3241; EGR4.
DR   MIM; 128992;
DR   GO; GO:0008284; P:positive regulation of cell proliferation; TAS.
DR   InterPro; IPR007087; ZnF_C2H2.
DR   Pfam; PF00096; zf-C2H2; 3.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 3.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
KW   Nuclear protein; Transcription regulation; DNA-binding; Zinc-finger;
KW   Metal-binding; Repeat.
FT   ZN_FING     380    404       C2H2-TYPE 1.
FT   ZN_FING     410    432       C2H2-TYPE 2.
FT   ZN_FING     438    460       C2H2-TYPE 3.
FT   CONFLICT    427    427       S -> T (IN REF. 2).
SQ   SEQUENCE   486 AA;  50855 MW;  0DF764427B0A21E3 CRC64;

Query Match             6.6%;  Score 82;  DB 1;  Length 486;
Best Local Similarity   24.6%;  Pred. No. 7.9;
Matches   42;  Conservative   23;  Mismatches   58;  Indels   48;  Gaps    9;
```

```
Qy      1 PWRADDILAGPRRLLDPQ----PYP--------GAPHHGSYVHFQP--------A 35
             |||   |  ||||         |||               | |
Db    131 PAGSDALLFGPPDLYSPDLGAAPPEAFWEASPCAGAP--SQCLYEPQLSPPDVKPGLR 187

Qy     36 RPTGGFVHTHTHTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAF 95
             |||    |                 |:::  :|||
Db    188 RPPASPALDAVSAFKGPYAPWELLSVGAPGNCGSQG----DYQAAPRARF--------PV 235

Qy     96 LSSRLQDLYSIVRRADRTGPVVNLRDEVLFPS--WEALFSGSEGQLKPGA 144
             :::::||                  |:| |
Db    236 IGTKIEDLLSCPAELPAVPA-----NRLYPSGAYDA-FPLAPGDLGEGA 280
```

RESULT 14
SR13_HUMAN
```
ID   SR13_HUMAN     STANDARD;      PRT;   995 AA.
AC   Q9Y3M8;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   StAR-related lipid transfer protein 13 (StARD13) (START domain-
DE   containing protein 13) (46H23.2).
GN   STARD13 OR GT650.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rhodes S.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Hunt A.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
```
```
DR   EMBL; AL049801; CAB42562.1; -.
DR   EMBL; Z84483; CAC94774.1; -.
DR   PIR; H59432; H59432.
DR   Genew; HGNC:19164; STARD13.
DR   InterPro; IPR008936; Rho_GAP.
DR   InterPro; IPR000198; RhoGAP.
DR   InterPro; IPR002913; START.
DR   Pfam; PF00620; RhoGAP; 1.
DR   Pfam; PF01852; START; 1.
DR   SMART; SM00324; RhoGAP; 1.
DR   SMART; SM00234; START; 1.
DR   PROSITE; PS50238; RHOGAP; 1.
DR   PROSITE; PS50848; START; 1.
KW   GTPase activation.
FT   DOMAIN      545    750       RHO-GAP.
FT   DOMAIN      781    989       START.
SQ   SEQUENCE   995 AA;  111191 MW;  3F608FA94A4BF8BF CRC64;

Query Match             6.6%;  Score 82;  DB 1;  Length 995;
Best Local Similarity   23.0%;  Pred. No. 18;
Matches   37;  Conservative   26;  Mismatches   52;  Indels   46;  Gaps    7;
```

```
Qy     98 SRLQDLYSIVRRADRTGVP------------VVNLRDEVLFPSWEALFSGSEGQLKPG 143
             ||:||| :|  |:| ||                 :|||||||||||||||||
Db     36 SRVDDLYTLLPRGDRNGSPGGTGMRNTTSSESVLTDLSBPEVCSIHSESSGGSDSRSQPG 95
```

Qy   144 --------ARIFSPD----GRDVLQHPAWPRKSVWHGSDPSGRRLTDSY---CETWR 185
          | ||| || ||||||||||||||||| | ||||
Db   96 QCCTDNPVMLDAPLVSSSLPQPPRDVLNHPFHPK-----NEKPTRARAKSFLKRMETLR 149

Qy   186 TEAPAATGQASS-----LLAGRLLEQEAASCRHAFVVLCIE 221
          : :   ::: |:|:||| |||
Db   150 GKGAHGRHKGSGRTGGLVISGPMLQQEPESFK---AMQCIQ 187

RESULT 15
NRT2_CHICK
ID   NRT2_CHICK   STANDARD;   PRT;   312 AA.
AC   P55807;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   NAD(P)(+)--arginine ADP-ribosyltransferase 2 precursor (EC 2.4.2.31)
DE   (Mono(ADP-ribosyl)transferase 2) (AT2).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=White leghorn; TISSUE=Bone marrow;
RX   MEDLINE=95050487; PubMed=7961658;
RA   Tsuchiya M., Hara N., Yamada K., Osago H., Shimoyama M.;
RT   "Cloning and expression of cDNA for arginine-specific ADP-
RT   ribosyltransferase from chicken bone marrow cells.";
RL   J. Biol. Chem. 269:27451-27457(1994).
CC   -!- CATALYTIC ACTIVITY: NAD(+) + L-arginine = nicotinamide + N(2)-
CC   (ADP-D-ribosyl)-L-arginine.
CC   -!- SUBCELLULAR LOCATION: THE MATURE ENZYME IS PROBABLY SECRETED
CC   FROM GRANULOCYTES INTO THE EXTRACELLULAR SPACE.
CC   -!- SIMILARITY: Belongs to the Arg-specific ADP-ribosyltransferase
CC   family.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; D31865; BAA06665.1; -.
DR   PIR; B55461; B55461.
DR   InterPro; IPR000768; ART.
DR   Pfam; PF01129; ART; 1.
DR   PRINTS; PR00970; RIBTRNSFRASE.
DR   PROSITE; PS01291; ART; 1.
KW   Transferase; Glycosyltransferase; NAD; Signal; Zymogen.
FT   SIGNAL   1   20   POTENTIAL.
FT   PROPEP   21   31   BY SIMILARITY.
FT   CHAIN   32   266   NAD(P)(+)--ARGININE ADP-
FT   RIBOSYLTRANSFERASE 2.
FT   PROPEP   267   312   POTENTIAL.
FT   ACT_SITE   231   231   BY SIMILARITY.
SQ   SEQUENCE   312 AA; 34966 MW; 9E45A9B7A6BC4057 CRC64;

Query Match   6.6%; Score 81.5; DB 1; Length 312;
Best Local Similarity 26.5%; Pred. No. 5.3;
  Matches   53; Conservative 16; Mismatches 68; Indels 63; Gaps 13;

Qy   22 GAPIHGSYVHFQPARPTGGPVHTHTHTHQDFQLVLHLVALNSPQP------GGMRGIRGA 75
          |||||| |:|||||:|||:|:|:| |:|||||: |:||||
Db   125 GGCSHQYYVHFY---------HFKT-LHFLLTQALFALRASQPRCYVYVYRGVRGIR-- 170

Qy   76 DFQCFQQARAAGLAGTFRAFLSSRL--------QDLYSIVRRADRTGVPVVNL----R 121
           : : ||:|| :   ::: ||: ::| :::
Db   171 ----FMTQR--GKSVRFGQFTSTSLRKDVAVNFGQDTFFVKTC--YGVPIKQFSFYFSE 222

Qy   122 DEVLFPSWEAL----FSGSEGQLKPGARI---FSFDGRDVLQHPAWPRKSVW-HGSDPSG 173
          ||||||||| |:|||||||||||| |:||||||||||||||||| |||||||
Db   223 DEVLIPPEVFEVTNFCTGNGRIQIYLRSKGKOMSRHNCELLK----PRGGQWGRGHQEVG 278

Qy   174 RRLTD--------SYCETW 184
          ||||:   :||||||
Db   279 LGLSPGLALPVLPCSNCSCW 298

Search completed: March 26, 2004, 13:36:31
Job time : 15.6667 secs

```
              GenCore version 5.1.6
       Copyright  (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:    March 26, 2004, 13:32:37 ; Search time 40.5556 Seconds
                                        (without alignments)
                                       1789.377 Million cell updates/sec

Title:          US-09-938-391-2
Perfect score:  1239
Sequence:       1 PWRADDILAGPPRLLDPQPY.........CRHAFVLCIENSVMTSFSK 230

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:    1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      SPTREMBL_25:*
           1:   sp_archea:*
           2:   sp_bacteria:*
           3:   sp_fungi:*
           4:   sp_human:*
           5:   sp_invertebrate:*
           6:   sp_mammal:*
           7:   sp_mhc:*
           8:   sp_organelle:*
           9:   sp_phage:*
          10:   sp_plant:*
          11:   sp_rodent:*
          12:   sp_virus:*
          13:   sp_vertebrate:*
          14:   sp_unclassified:*
          15:   sp_rvirus:*
          16:   sp_bacteriap:*
          17:   sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                      SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1019	82.2	816	4	Q8N4S4	Q8n4s4 homo sapien
2	983.5	79.4	1140	11	Q61434	Q61434 mus musculu
3	897.5	72.4	226	12	Q9Q2D2	Q9qd2 rattus norv
4	837	67.6	261	4	Q8NG19	Q8ng19 homo sapien
5	835	67.4	187	4	Q8WX15	Q8wx15 homo sapien
6	792	63.9	1344	13	O93419	O93419 gallus gall
7	754	60.9	171	11	Q9WUW5	Q9wuw5 rattus norv
8	736.5	59.4	1315	13	Q8QHL9	Q8qhl9 xenopus lae
9	730	58.9	1307	13	Q8JFF7	Q8jff7 xenopus lae
10	726	58.6	160	11	Q9CRT2	Q9crt2 mus musculu
11	658.5	53.1	361	13	Q8AMC6	Q8awc6 brachydanio
12	562.5	45.4	1367	11	O35206	O35206 mus musculu
13	561.5	45.3	1367	11	Q9EQD9	Q9eqd9 mus musculu
14	527	42.5	1388	4	Q9Y4W4	Q9y4w4 homo sapien
15	457.5	36.9	950	5	Q86SC8	Q86sc8 ciona intes
16	456	36.8	102	4	Q96T70	Q96t70 homo sapien

```
17   446.5  36.0   792   5   Q8MT89   Q8mt89 drosophila
18   434.5  35.1   299   5   Q8MSE3   Q8mse3 drosophila
19   428.5  34.6   778   5   Q86BH1   Q86bh1 drosophila
20   369.5  29.8   650   5   O17866   O17866 caenorhabdi
21   369.5  29.8   778   5   Q9U9K6   Q9u9k6 caenorhabdi
22   369.5  29.8  1117   5   Q9U9K7   Q9u9k7 caenorhabdi
23   369.5  29.8  1154   5   Q8IOG6   Q8iOg6 caenorhabdi
24   101.5   8.2   287  16   Q8JI38   Q8ji38 bradyrhizob
25   100.5   8.1   739  16   Q82BY9   Q82by9 streptomyce
26   100.5   8.1   745  16   Q8P380   Q8p380 xanthomonas
27    99     8.0   244   2   Q53235   Q53235 rhodobacter
28    99     8.0   521  15   Q72L07   Q72l07 recombinant
29    98     7.9   427  15   Q62700   Q62700 rattus norv
30    98     7.9   521  15   Q72L05   Q72l05 recombinant
31    98     7.9  2301  10   Q9ATK5   Q9atk5 chlamydomon
32    97.5   7.9   651   5   Q9VFA9   Q9vfa9 drosophila
33    94.5   7.6   208  16   Q92K28   Q92k28 rhizobium m
34    94     7.6   830  16   Q8CJK2   Q8cjk2 streptomyce
35    93.5   7.5   702   2   O52660   O52660 escherichia
36    93.5   7.5   713   2   O52679   O52679 escherichia
37    93     7.5   517  15   Q83378   Q83378 rat leukemi
38    91.5   7.4  1895   5   Q9VW95   Q9vw95 drosophila
39    91     7.3   439  10   Q8LIQ7   Q8liq7 oryza sativ
40    91     7.3  6239  16   Q9S0R7   Q9s0r7 streptomyce
41    90     7.3   404   9   Q9XJK3   Q9xjk3 bacteriopha
42    90     7.3   404   4   Q7Y2X3   Q7y2x3 stx2 conver
43    90     7.3   404  16   Q9KXC4   Q9kxc4 escherichia
44    90     7.3   776  16   Q88E70   Q88e70 pseudomonas
45    89     7.2   318   9   Q8SCB2   Q8scb2 stx2 conver
```

```
                            ALIGNMENTS

RESULT 1
Q8N4S4
ID   Q8N4S4      PRELIMINARY;      PRT;   816 AA.
AC   Q8N4S4;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Similar to collagen, type XVIII, alpha 1 (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Renal adenocarcinoma;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC033715; AAH33715.1; -.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   Pfam; PF01391; Collagen; 5.
DR   ProDom; PD000007; Clg_helix; 1.
KW   Collagen.
FT   NON_TER       1          1
SQ   SEQUENCE    816 AA;   82553 MW;   5D539B294694F86 CRC64;

Query Match              82.2%;    Score 1019;  DB 4;  Length 816;
Best Local Similarity    82.4%;    Pred. No. 1.8e-89;
Matches  187;  Conservative  16;  Mismatches  20;  Indels  4;  Gaps  1;

Qy     1 PWRADDILAGPPRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHTHQDFQLVLHLVA   60
          |||||||||||||||||||  |||||:| ||:||||||||:|||||||:||||||  ||||
Db   592 PWRADDILASPPRLPEPQYPGAPHHSSYVHLRPARPTSPP----AHSHRDFQPVLHLVA  647

Qy    61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL  120
          |||:|.||||||||||||||||||:||||||||||||||||||||||| |.|||:||||:
Db   648 LNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAVPIVNL  707

Qy   121 RDEVLFPSWEALFSSSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWEGSDPSGRRLTDSY  180
```

```
Db   708  :||:||||||||||  ||||||||||  |:||  |:|  ||:||  :||||||||:||
          KDELLFPSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPQKSVWHGSDPNGRRLTBSY 767

Qy   181  CETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTS 227
          |||||||  ||||||||||||  |||:|||  |:  ||:||||||  ||:
Db   768  CETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 814


RESULT 2
Q61434
ID   Q61434      PRELIMINARY;       PRT;   1140 AA.
AC   Q61434;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Collagen (Fragment).
GN   COL15A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Abe N., Muragaki Y., Yoshioka H., Inoue H., Ninomiya Y.;
RT   "Identification of a novel collagen chain represented by extensive
RT   interruptions in the triple-helical region.";
RL   Cell. Mol. Biol. Res. 196:576-582(1993).
DR   EMBL; D17546; BAA04483.1; -.
DR   PIR; B56101; B56101.
DR   HSSP; P39061; 1KOE.
DR   MGD; MGI:88449; Col15a1.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 8.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
KW   Collagen.
FT   NON_TER      1      1
SQ   SEQUENCE   1140 AA;  115156 MW;  8B0C7E6862B3BDFE CRC64;

Query Match             79.4%;  Score 983.5;  DB 11;  Pred. No. 7.4e-86;
Best Local Similarity   80.1%;
Matches  185;  Conservative  18;  Mismatches  23;  Indels  5;  Gaps  2;

Qy     1  PWRADDILAGPPRLLDPQPYPGAP-HHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLV 59
          ||||||  |:|||||||||||||||  :||||||  ||||||||||||||||||||||||
Db   914  PWRADDILANPPRLLPDRQPYPGVPHHHSSYVHLPPARPT----LSLAHTHQDFQPVLHLV 969

Qy    60  ALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVN 119
          |||| ||||||||||||||||||||||:||||||||||||||||||||||||:|:|||
Db   970  ALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIVN 1029

Qy   120  LRDEVLPPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDS 179
          |:|||:||:||||||  |||:|||||||||||||||:|||||||||||||||||:
Db  1030  LKDEVLSPSWNDSLFSGSGGQLQPGARIFSFDGRDVLRHPAWPQKSVWEGSDPSGRRLMES 1089

Qy   180  YCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 230
          |||||||||  |||||||||  |||||||||:::||||||||||||||||||
Db  1090  YCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMTSFSK 1140


RESULT 3
Q9Q2D2
ID   Q9Q2D2      PRELIMINARY;       PRT;   226 AA.
AC   Q9Q2D2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Collagen XVIII (Fragment).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=20227226; PubMed=10766159;
RA   Perletti G., Concari P., Giardini R., Marras E., Piccinini F.,
RA   Folkman J., Chen L.;
RT   "Antitumor activity of endostatin against carcinogen-induced rat
RT   primary mammary tumors.";
RL   Cancer Res. 50:1793-1796(2000).
DR   EMBL; AF189709; AAF00975.1; -.
DR   HSSP; P39061; 1KOE.
FT   NON_TER      1      1
SQ   SEQUENCE    226 AA;  25350 MW;  38B83C0486C0E949 CRC64;

Query Match             72.4%;  Score 897.5;  DB 11;  Length 226;
Best Local Similarity   74.2%;  Pred. No. 1.8e-78;
Matches  170;  Conservative  18;  Mismatches  28;  Indels  13;  Gaps  2;

Qy    11  PPRLLDPQPYPGAP--------HHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLVAL 61
          ||||  |||:||||            :||||||||||||||||||||||||||||:||
Db     2  PPRLPDRQPYPGVPHHHHHHHHHHSSHEHRPPAHPSPSP---AHTHQDFHPVLHLVAL 57

Qy    62  NSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNLR 121
          :||  |||||||||||||||||||:|||||||||||||||||||||||  :|::||
Db    58  NTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRSSVPIVNLK 117

Qy   122  DEVLFPSWEALFSGSSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSYC 181
          |||| ||| ||:|||  |||:|||||||||||||:|||:||||:||||||||||:::|||
Db   118  DEVLSPSWDTLFSGSQGQLHSGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLMESYC 177

Qy   182  ETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 230
          ||||||||  |||||||||  |||||||||:::|||||||||||||||||
Db   178  ETWRTEATGVTGQASSLLSGRLLEQKAESCHNSYIVLCIENSFMTSFSK 226


RESULT 4
Q8NG19
ID   Q8NG19      PRELIMINARY;       PRT;   261 AA.
AC   Q8NG19;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Multi-functional protein MFP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Dou D.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF282883; AAM52249.1; -.
SQ   SEQUENCE    261 AA;  26745 MW;  CA60C920AF3B90B5 CRC64;

Query Match             67.6%;  Score 837;  DB 4;  Length 261;
Best Local Similarity   75.1%;  Pred. No. 1.5e-72;
Matches  163;  Conservative  16;  Mismatches  24;  Indels  14;  Gaps  3;

Qy    11  PPRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHQDFQLVLHLVALNSPQPGGMR 70
          ::||||  :||      :::   ||||:           |||||||||||:|:|||||:
Db    57  PRKLYD---YGGGSGGGS-----GHSHRDFQPVLHLVALNSPLSGGMR 102

Qy    71  GIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNLRDEVLFPSWE 130
          |||||||||||||||:|||:|||||||||||||||||||||||:: |::
Db   103  GIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRAAVPIVNLKDELLFPSWE 162

Qy   131  ALFSGGEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSYCETWRTEAPA 190
          ||||| ||||:|||||||||||||:|||:||||:|||||||| :|::||||||||:
Db   163  ALFSGSEGPLKPGARIFSFDGDKVLRHPTWPQKSVWGSDPNGRRLTESYCETWRTEAPS 222
```

```
Qy    191 ATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTS 227
          ||||||||||||||||||||| ||::|||||||| ||
Db    223 ATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTA 259


RESULT 5
Q8WXI5
ID   Q8WXI5            PRELIMINARY;     PRT;    187 AA.
AC   Q8WXI5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Collagen XVIII (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21409408; PubMed=11517600;
RA   Feng Y., Cui L.B., Liu C.X., Ma Q.J.;
RT   "Inhibition effect in vitro of purified endostatin expressed in Pichia
RT   pastoris.";
RL   Sheng Wu Gong Cheng Xue Bao 17:278-282(2001).
DR   EMBL; AF416592; AAL37720.1; -.
FT   NON_TER      1      1
SQ   SEQUENCE   187 AA;  20448 MW;  72B1047D85838CD3 CRC64;

Query Match             67.4%;  Score 835; DB 4; Length 187;
Best Local Similarity   85.1%;  Pred. No. 1.6e-72;
Matches  154;  Conservative  14;  Mismatches  13;  Indels  0;  Gaps  0;

Qy    47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106
         :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 64

Qy   107 VRRADRTGVPVVNLRDEVLFPSWEALFGSSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 166
         ||||||:  |:||:||||||||||||||||||:||||||||||||||||||||||||||
Db    65 VRRADRAAVPIVNLKDELLFPSWEALFGSSEGPIKPGARIFSFDGKDVLRHPTWPQKSVW 124

Qy   167 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 226
         ||:|:|||| :||||||||||| |||||||||||||||||||||| |||||||||||||
Db   125 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 184

Qy   227 S 227
         :
Db   185 A 185


RESULT 6
O93419
ID   O93419            PRELIMINARY;     PRT;   1344 AA.
AC   O93419;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Collagen XVIII precursor.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98411346; PubMed=9738008;
RA   Halfter W., Dong S., Schurer B., Cole G.J.;
RT   "Collagen XVIII is a basement membrane heparan sulfate proteoglycan.";
RL   J. Biol. Chem. 273:25404-25412(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Halfter W., Dong S.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF083440; AAC33294.2; -.
```

```
DR   HSSP; P39061; 1KOE.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008161; C1q.helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 8.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; C1q_helix; 2.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen; Signal.
FT   SIGNAL       1     26       POTENTIAL.
SQ   SEQUENCE   1344 AA;  137402 MW;  7AA366E4FE940CCD CRC64;

Query Match             63.9%;  Score 792; DB 13;  Length 1344;
Best Local Similarity   66.2%;  Pred. No. 2.7e-67;
Matches  155;  Conservative  20;  Mismatches  47;  Indels  12;  Gaps  4;

Qy     1 PWRADDILAGPPRLLDPQPYPGAPH------GSYVHFQPARPTGGPVHTHTHQDFQLVL 56
         ||  :|:  :|:|||       :: |
Db  1119 PWRGDEVVANQHLL--PQP-PLLQQHELLNSYIHRWP-----DPAPVAAHVHQDFQPAL 1170

Qy    57 HLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVP 116
         ||||||:|:|||||||||||||||||||| |:|||||||||||||||||||||||||:||
Db  1171 HLVALNTPLSGGMRGIRGADFQCFQQARQVGLAGTFRAFLSSRLQDLYSIVRRADRTAVP 1230

Qy   117 VVNLRDEVLFPSWEALFGSGEGQLKPGARIFSFDGRDVLQHPAMPRKSVWHGSDPSGRRL 176
         :|||||||||||||||||: :|||:||||||||||||||::||:  ::|||||||||:|||
Db  1231 IVNLRDEVLFSNWEALFTGSEAPLRAGARIISFDGRDILQDSAWPQKSIWHGSDAKGRRL 1290

Qy   177 TDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 230
         |||||||||||||:||||||||||||||||:|||::|||| ||||||||  ||
Db  1291 PESYCEAWRTDERGTSGQASSLSSGKLLEQSASSCHAFVVLCIENSFMTAAKK 1344


RESULT 7
Q9WUW5
ID   Q9WUW5            PRELIMINARY;     PRT;    171 AA.
AC   Q9WUW5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Collagen type XVIII, alpha (I) chain (Fragment).
GN   COL18A1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=liver;
RA   Jia J.D., Bauer M., Eberspaecher U., Donner P., Schuppan D.;
RT   "Partial 3'-terminal cDNA sequence of rat collagen XVIII/endostatin.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=liver;
RA   Jia J.D., Bauer M., Sedlaczek N., Ruehl M., Riecken E.O., Schuppan D.;
RT   "Temporospatial expression of collagen XVIII/endostatin in acute and
RT   chronic liver injuries.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ236873; CAB44263.1; -.
DR   HSSP; P39061; 1KOE.
FT   NON_TER      1      1
FT   NON_TER    171    171
SQ   SEQUENCE   171 AA;  18933 MW;  81BE2EB3FC2C8E72 CRC64;

Query Match             60.9%;  Score 754; DB 11;  Length 171;
Best Local Similarity   81.9%;  Pred. No. 9.1e-65;
Matches  140;  Conservative  16;  Mismatches  15;  Indels  0;  Gaps  0;
```

Qy   55 VLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTG 114
          |||||||:|:| |||||||||||||||||| :|:||||||||||:||||||||||||||
Db   1 VLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSGRLLEQKABSCHNSYIVLCIENSFM 60

Qy   115 VPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGR 174
          ||:|||:|||| ||||:|||||| |||: |||||||||||||:||||||:||||||||||
Db   61 VPIVNLKDEVLSPSWDTLFSGSGGQLHSGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGR 120

Qy   175 RLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 225
          || |:||||||||| |||||||||:|||||| ::::| ::||||||||||
Db   121 RLMESVCETWRTEATGVTGQASSLLSGRLLEQKABSCHNSYIVLCIENSFM 171

RESULT 8
Q8QHL9
ID   Q8QHL9   PRELIMINARY;   PRT;   1315 AA.
AC   Q8QHL9;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Type XVIII collagen alpha1 chain.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ishino T., Sekimizu K., Natori S., Kubo T.;
RT   "Idenfication and characterization of genes expressed selectively in
RT   the regenerating tail of Xenopus laevis tadpole.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB047066; BAB84674.1; -.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 7.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen.
SQ   SEQUENCE   1315 AA; 134946 MW; 0C56C235DB058365 CRC64;

Query Match   59.4%;
Best Local Similarity   62.6%; Pred. No. 6e-62; DB 13; Length 1315;
Matches 144; Conservative 24; Mismatches 55; Indels 7; Gaps 3;

Qy   1 PWRADDILAGPPRLLDDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHTHQDFQLVLHLVA 60
          |||:.:|| |: |: || : ...| .:| :|| |:| ||||||
Db   1093 PW-SDPRLPEPPRVDPPRVDPPPDHHHHGNPHPLEPS--PV----HTHQQFNPALHLVA 1145

Qy   61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
          |:.|| |||||:|||||||||::|| :||:||||||||||||||||||||||:.:.:|||
Db   1146 LNAPLSGGMKSIRGVDFQCFEQARKSGLHGTFRAFLSSRLQDLYSIVRRADRQSVQIVNL 1205

Qy   121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180
          |||||:..|::||||||.|::.||||||||||:||.:|:||:|.||||||:.|||||:||
Db   1206 RDEVLYDNWESLFSGSEAQMRPGARIFSFDGKDVATDPTWPQKMVWHGSDAKGRRLTESY 1265

Qy   181 CETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 230
          ||||||.::..|.||||| .|:|||||:..||::.|:|||||||:|...|
Db   1266 CETWRTDESAVTSQASSLTSGKLLEQRPQSCNKNFIVLCIENSFMTNNRK 1315

RESULT 9
Q8JFF7
ID   Q8JFF7   PRELIMINARY;   PRT;   1307 AA.
AC   Q8JFF7;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Type XVIII collagen short variant.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22166979; PubMed=12175494;
RA   Elamaa H., Peterson J., Pihlajaniemi T., Destree O.;
RT   "Cloning of three variants of type XVIII collagen and their expression
RT   patterns during Xenopus laevis development.";
RL   Mech. Dev. 114:109-113 (2002).
DR   EMBL; AY052763; AAL14257.1; -.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 6.
DR   Pfam; PF02210; TSPN; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen.
SQ   SEQUENCE   1307 AA; 134153 MW; D53EDBFE3DE34976 CRC64;

Query Match   58.9%;
Best Local Similarity   63.0%; Pred. No. 2.5e-61; DB 13; Length 1307;
Matches 145; Conservative 24; Mismatches 49; Indels 12; Gaps 5;

Qy   1 PWRADDILAGPPRLLDDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHTHQDFQLVLHLVA 60
          |||:.:|| |: |: || : ...| .:| :|| |:| ||||||
Db   1090 PW-SDPRLPEPPR-VDPPP----PDHHQHGNPRPLEPS--PV----HTHQDFNPALHLVA 1137

Qy   61 LNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNL 120
          |:.|| |||||:|||||||||::|| :||:||||||||||||||||||||||:.:.:|||
Db   1138 LNAPISGSMKSIRGVDFQCFEQARKAGLHGTFRAFLSSRLQDLYSIVRRADRQSVQIVNL 1197

Qy   121 RDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSY 180
          |||||:..|::||||||.|::.||||.||||:||..:|:||:|.||||||:.|||||:||
Db   1198 RDEVLYDNWDSLFSGSEAQMRSGARILSFDGKDVTTDPTWPQKMVWHGSDAKGRRLTESY 1257

Qy   181 CETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 230
          ||||||.::..|.||||| .|:|||||:..||::.|:|||||||:|..||
Db   1258 CETWRTDESAVTGQASSLTGKLLEQRPQSCNKNFIVLCIENSFMTNNRK 1307

RESULT 10
Q9CRT2
ID   Q9CRT2   PRELIMINARY;   PRT;   160 AA.
AC   Q9CRT2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Procollagen, type XVIII, alpha 1 (Fragment).
GN   COL18A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa H., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

```
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK014292; BAB22449.1; -.
DR   HSSP; P39061; 1KOE.
DR   MGD; MGI:88451; Col18a1.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   GO; GO:0001525; P:angiogenesis; IMP.
FT   NON_TER       1      1
SQ   SEQUENCE   160 AA;  17725 MW;  60F853D777C375D2 CRC64;

Query Match             58.6%;  Score 726;  DB 11;  Length 160;
Best Local Similarity   83.8%;  Pred. No. 4.2e-62;  -.
Matches  134;  Conservative   16;  Mismatches   10;  Indels    0;  Gaps    0;

Qy     71 GIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNLRDEVLFPSWE 130
          |||||||||||||||| ||:|||||||||||||||||||||| ||:|||:|||||:|||:
Db      1 GIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIVNLKDEVLSPSWD 60

Qy    131 ALFSGSSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSYCETWRTEAPA 190
           :|||| ||:|||||||||||||||| :|||| |||||||||||||| :|||||||||:
Db     61 SLFSGSQGLQPGARIFSFDGRDVLRIPAWFQKSVWHGSDPSGRRLMESYCETWRTETTG 120

Qy    191 ATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 230
          ||||||||:|||||||:::|||:|::|||||||:|||||||
Db    121 ATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFWTSFSK 160

RESULT 11
ID   Q8AWC6
AC   Q8AWC6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Collagen XVIII (Fragment)
GN   COL18A1.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Haftek Z., Morvan-Dubois G., Thisse B., Garrone R., Le Guellec D.;
RT   "Sequence and embryonic expression of collagen XVIII NC11 domain
RT   (endostatin) in the zebrafish.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ494837; CAD38825.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   361 AA;  40222 MW;  3C5A0F8479D26735 CRC64;

Query Match             53.1%;  Score 658.5;  DB 13;  Length 361;
Best Local Similarity   58.4%;  Pred. No. 3.9e-55;
Matches  136;  Conservative   23;  Mismatches   55;  Indels   19;  Gaps    6;

Qy     11 PP----RLLDP-QP--YPGAPHHGSYVHFQPARPTGGPV----HTHTHTHQDFQLVLH 57
          ||     ::||| ||   ||:||||||||| |||: :    |:| ||::|||||||
Db    135 PFYTDNRYTDPVEPNIRYPVQPERNP---ITPARRPIPPVNQPCGHAHTHTSGP---GLH 188

Qy     58 LVALNSPQPGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPV 117
          |:|||||| |:|||||||||:|||||||:|:.|||||||| :||||||||||:|| |
Db    189 LIALNSPQVGNMRGIRGADFMCFQQARAVGMKGTFRAFLSSKIQDLYSIVRRSDRETLPI 248

Qy    118 VNLRDEVLFPSWEALFSGSSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLT 177
          |||:|:||| ||: ||| : :: :  :||||||||||||| :||| ::||||||: |:|:|
```

```
Db    249 VNLKDQVLFSSWESLFSDSESRMKDNAPIYSFDGRDVLRDSAWPEKMIWHGSDGRGHRQT 308
Qy    178 DSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 230
          |:|||||| |   .||| |||| |||| ||::||   ::|||||||:||: | 
Db    309 DNYCETWRAGDRAVTGLASSLQAGLLQQTSSSCSSSYIALCIENSVMTQSKK 361

RESULT 12
ID   O35206
AC   O35206;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type XV collagen.
GN   COL15A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97480713; PubMed=9339358;
RA   Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA   Pihlajaniemi T.;
RT   "Cloning of mouse type XV collagen sequences and mapping of the
RT   corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
RT   (XV) collagen sequences indicates divergence in the number of small
RT   collagenous domains.";
RL   Genomics 45:31-41(1997).
DR   EMBL; AF011450; AAC53387.1; -.
DR   PDB; 1DY2; 21-JAN-01.
DR   MGD; MGI:88449; Col15a1.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008161; Clg.helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 5.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen.
SQ   SEQUENCE   1367 AA;  140525 MW;  A483A1254AF3AEEC CRC64;

Query Match             45.4%;  Score 562.5;  DB 11;  Length 1367;
Best Local Similarity   51.4%;  Pred. No. 3.9e-45;
Matches  113;  Conservative   26;  Mismatches   60;  Indels   21;  Gaps    2;

Qy     11 PPRLLDPQPYPGAPHHGSYVHFQPARPTGGPVHTHTHTHQDFQLVLHLVALNSPQPGGMR 70
          ||   : :||  |  ||||||:|||: ||||:             :||||||||:||  :
Db   1169 PPPALSSNPYQPQPPLNPILSANYERP------------VLHLVALNTPVAGDIR 1211

Qy     71 GIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNLRDEVLFPSWE 130
          :  |||||||||||||||| .||||||| :|||| ::|| |: | |:|||: ::||  :|
Db   1212 ---ADFQCFQQARAAGLLSTFRAFLSSHLQDLSTVVRKAERFGLPIVNLKGQVLFNNWD 1267

Qy    131 ALFSGSSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSYCETWRTEAPA 190
          ::||| . : :  .|  |:|||||||:  :|| :||||||:  ||||: |||| |||:: |
Db   1268 SIFSGDGQFNTHIPIYSFDGRDVMTDPSWFQKVVWHGSNHGVRVLVDKYCEAWRTTDMA 1327

Qy    191 ATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 230
          .|| ||  |   ::| :| ||    ::||||||:|| : |
Db   1328 VTGFASPLSTGKILDQKAYSCANRLIVLCIENSFWTDTRK 1367

RESULT 13
ID   Q9EQD9
AC   Q9EQD9;
```

```
Db    249 VNLKDQVLFSSWESLFSDSESRMKDNAPIYSFDGRDVLRDSAWPEKMIWHGSDGRGHRQT 308
Qy    178 DSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 230
Db    309 DNYCETWRAGDRAVTGLASSLQAGLLQQTSSSCSSSYIALCIENSVMTQSKK 361
```

O35206   PRELIMINARY;   PRT;   1367 AA.

Q8AWC6   PRELIMINARY;   PRT;   361 AA.

Q9EQD9   PRELIMINARY;   PRT;   1367 AA.

```
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type XV collagen.
GN   COL15A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=97480713; PubMed=9339358;
RA   Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA   Pihlajaniemi T.;
RT   "Cloning of mouse type XV collagen sequences and mapping of the
RT   corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
RT   (XV) collagen sequences indicates divergence in the number of small
RT   collagenous domains.";
RL   Genomics 45:31-41(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=20522048; PubMed=11068203;
RA   Eklund L., Muona A., Lietard J., Pihlajaniemi T.;
RT   "Structure of the mouse type XV collagen gene, Col15a1, comparison
RT   with the human COL15A1 gene and functional analysis of the promoters
RT   of both genes.";
RL   Matrix Biol. 19:489-500(2000).
DR   EMBL; AF261131; AAG27545.1; JOINED.
DR   EMBL; AF261109; AAG27545.1; JOINED.
DR   EMBL; AF261110; AAG27545.1; JOINED.
DR   EMBL; AF261111; AAG27545.1; JOINED.
DR   EMBL; AF261112; AAG27545.1; JOINED.
DR   EMBL; AF261113; AAG27545.1; JOINED.
DR   EMBL; AF261114; AAG27545.1; JOINED.
DR   EMBL; AF261115; AAG27545.1; JOINED.
DR   EMBL; AF261116; AAG27545.1; JOINED.
DR   EMBL; AF261117; AAG27545.1; JOINED.
DR   EMBL; AF261118; AAG27545.1; JOINED.
DR   EMBL; AF261119; AAG27545.1; JOINED.
DR   EMBL; AF261120; AAG27545.1; JOINED.
DR   EMBL; AF261121; AAG27545.1; JOINED.
DR   EMBL; AF261122; AAG27545.1; JOINED.
DR   EMBL; AF261123; AAG27545.1; JOINED.
DR   EMBL; AF261124; AAG27545.1; JOINED.
DR   EMBL; AF261125; AAG27545.1; JOINED.
DR   EMBL; AF261126; AAG27545.1; JOINED.
DR   EMBL; AF261127; AAG27545.1; JOINED.
DR   EMBL; AF261128; AAG27545.1; JOINED.
DR   EMBL; AF261129; AAG27545.1; JOINED.
DR   EMBL; AF261130; AAG27545.1; JOINED.
DR   HSSP; P39061; 1KOE.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 5.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen.
SQ   SEQUENCE   1367 AA;  140454 MW;  40F259A3CD47C982 CRC64;

Query Match             45.3%;    Score 561.5;  DB 11;  Length 1367;
Best Local Similarity   51.4%;    Pred. No. 4.8e-45;
Matches 113;  Conservative 26;  Mismatches 60;  Indels 21;  Gaps  2;

Qy    11  PPRLLDPQPYPGAPHHGSVVHFQPARPTGGPVHTHTHTHQDFQLVLHLVALNSPQPGGWR  70
```

```
Db  1169  PPPALSSNPYQQPQPPLNPILSANYERP------------------VLHLVALNTPVAGDIR  1211
                                                         |||||||:||.|:||
Qy    71  GIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNLRDEVLFPSWE  130
          |:.||||||||||||||:.:|||||:||.:|||:|:|:|||:|:|..:||:.:|||..|:
Db  1212  ---ADFQCFQQARAAGLLSTFRAFISSHIQDLSTVVRKAERFGLPIVNLKGQVLFNNWD  1267
Qy   131  ALFSGSGQLKGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSYCETWRTEAPA  190
          :.:|||:||:.||:.|||||||:..|:||.:|.||||:|:|.||.|.||||:||.::|
Db  1268  SIFSGDGGQFNTHIPIYSFDGRDVMTDPSWPQKVVWHGSNPYGVRLVDKYCEAWRTTDMA  1327
Qy   191  ATGQASLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK  230
          .||.||:|:.|:|:|:.| |:|.:||||||||:||:|:|
Db  1328  VTGFASPLSTGKILDQKAYSCANRLIVLCIENSFMTDTRK  1367
```

```
RESULT 14
Q9Y4W4
ID   Q9Y4W4      PRELIMINARY;      PRT;   1388 AA.
AC   Q9Y4W4;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type XV collagen.
GN   COL15A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94148920; PubMed=8106446;
RA   Kivirikko S., Heinamaki P., Rehn M., Honkanen N., Myers J.C.,
RA   Pihlajaniemi T.;
RT   "Primary structure of the alpha 1 chain of human type XV collagen and
RT   exon-intron organization in the 3' region of the corresponding gene.";
RL   J. Biol. Chem. 269:4773-4779(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98316357; PubMed=9651385;
RA   Hagg P.M., Muona A., Lietard J., Kivirikko S., Pihlajaniemi T.;
RT   "Complete exon-intron organization of the human gene for the alpha1
RT   chain of type XV collagen (COL15A1) and comparison with the homologous
RT   COL18A1 gene.";
RL   J. Biol. Chem. 273:17824-17831(1998).
DR   EMBL; L25280; AAC78500.1; -.
DR   EMBL; AF052956; AAC78500.1; JOINED.
DR   EMBL; AF052957; AAC78500.1; JOINED.
DR   EMBL; AF052958; AAC78500.1; JOINED.
DR   EMBL; AF052959; AAC78500.1; JOINED.
DR   EMBL; AF052960; AAC78500.1; JOINED.
DR   EMBL; AF052961; AAC78500.1; JOINED.
DR   EMBL; AF052962; AAC78500.1; JOINED.
DR   EMBL; AF052963; AAC78500.1; JOINED.
DR   EMBL; AF052964; AAC78500.1; JOINED.
DR   EMBL; AF052965; AAC78500.1; JOINED.
DR   EMBL; AF052966; AAC78500.1; JOINED.
DR   EMBL; AF052967; AAC78500.1; JOINED.
DR   EMBL; AF052968; AAC78500.1; JOINED.
DR   EMBL; AF052969; AAC78500.1; JOINED.
DR   EMBL; AF052970; AAC78500.1; JOINED.
DR   EMBL; AF052971; AAC78500.1; JOINED.
DR   EMBL; AF052972; AAC78500.1; JOINED.
DR   EMBL; AF052973; AAC78500.1; JOINED.
DR   EMBL; AF052974; AAC78500.1; JOINED.
DR   EMBL; AF052975; AAC78500.1; JOINED.
DR   EMBL; L25285; AAC78500.1; JOINED.
DR   EMBL; L25284; AAC78500.1; JOINED.
DR   EMBL; L25283; AAC78500.1; JOINED.
DR   EMBL; L25282; AAC78500.1; JOINED.
DR   EMBL; L25281; AAC78500.1; JOINED.
DR   HSSP; P39061; 1KOE.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
```

Job time : 45.5556 secs

```
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 5.
DR   Pfam; PF02210; TSPN; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen.
SQ   SEQUENCE   1388 AA;   141757 MW;   9682845B847194B CRC64;

Query Match            42.5%;   Score 527; DB 4; Length 1388;
Best Local Similarity  56.6%;   Pred. No. 1e-41;
Matches  99;  Conservative  27;  Mismatches  45;  Indels   4;  Gaps   1;

Qy       56  LHIVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGV  115
                :|| ||    ::|||  || |:|:||||| |||    :|:|||||::||:|:|||
Db     1218  LHLAALNMPVSGDIR---ADFQCFKQARAAGLLSTYRAFLSSHLQDLSTIVRKAERYSL  1273

Qy      116  PVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRR  175
                |:||||||: :||:::|||  |||      ||||||| ::|:||  :|||||||| |
Db     1274  PIVNLKGQVLFNWDSIFSGHGGQFNMHIPIYSFDGRDIMTDPSWPQKVIWHGSSPHGVR  1333

Qy      176  LTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK  230
                |:|||| ||    ||||| ||:||:    :|:  ||:||||||||||||
Db     1334  LVDNYCEAWRTADTAVTGLASPLSTGKILDQKAYSCANRLIVLCIENSFMTDARK  1388


RESULT 15
Q86SC8
ID   Q86SC8          PRELIMINARY;           PRT;    950 AA.
AC   Q86SC8;
DT   01-JUN-2003  (TrEMBLrel. 24, Created)
DT   01-JUN-2003  (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003  (TrEMBLrel. 25, Last annotation update)
DE   Collagen XVIII homologue.
GN   CI-COLLAGEN XVIII.
OS   Ciona intestinalis.
OC   Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC   Phlebobranchia; Cionidae; Ciona.
OX   NCBI_TaxID=7719;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hotta K., Takahashi H., Satoh N.;
RT   "Further Characterization of Brachyury-Downstream Genes in Ciona
RT   intestinalis Embryo.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB076898; BAC57521.1; -.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001759; Pentaxin.
DR   Pfam; PF01391; Collagen; 8.
DR   PROSITE; PS00289; PENTAXIN; 1.
SQ   SEQUENCE   950 AA;   94768 MW;   2767ECA158CBA89F CRC64;

Query Match            36.9%;   Score 457.5; DB 5; Length 950;
Best Local Similarity  55.2%;   Pred. No. 3.3e-35;
Matches  96;  Conservative  17;  Mismatches  60;  Indels   1;  Gaps   1;

Qy       53  QLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADR  112
                :|:|||||:||||||||||||||||||||| ||:|||||| |:::|||||||||||
Db      773  QKSLHMIAINFPLRGNTRGIVGADARCFQQARRAGLKGTYRAFLSSRDQNVRSIVRREDR  832

Qy      113  TGVPVVNLRDEVLFPSWEALFSSSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPS  172
                | || ||||||:|||::|||||| :||||||||||||||||||||:||:|||||| |
Db      833  RNVPIVNIRGEQLFSSWEELFR-TEGRMDNPNMIYSFENRQVSTDARWPVKFVWHGSYTD  891

Qy      173  GRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT  226
                 ||:|:|||| :||   | ||| |||:||||||:|:|  ||:||||||||||
Db      892  GRLNPMHYCASWYTDHKAVTGQASPLSTRRLLAQKPYSCESGFVLCVENSTRT  945
```

Search completed: March 26, 2004, 13:37:58

OM nucleic - nucleic search, using sw model

Copyright GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: March 29, 2004, 09:11:19 ; Search time 2233.23 Seconds
                                            (without alignments)
                                            10771.549 Million cell updates/sec

Title: US-09-938-391-3
Perfect score: 555
Sequence: 1 cacacccaccaggacttcca........tgacctccttctccaagtag 555

Scoring table: IDENTITY NUC
               Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : GenEmbl:*
    1: gb_ba:*
    2: gb_htg:*
    3: gb_in:*
    4: gb_om:*
    5: gb_ov:*
    6: gb_pat:*
    7: gb_ph:*
    8: gb_pl:*
    9: gb_pr:*
   10: gb_ro:*
   11: gb_rod:*
   12: gb_sy:*
   13: gb_un:*
   14: gb_vi:*
   15: em_ba:*
   16: em_fun:*
   17: em_hum:*
   18: em_in:*
   19: em_mu:*
   20: em_om:*
   21: em_or:*
   22: em_ov:*
   23: em_pat:*
   24: em_ph:*
   25: em_pl:*
   26: em_ro:*
   27: em_sts:*
   28: em_un:*
   29: em_vi:*
   30: em_htg_hum:*
   31: em_htg_inv:*
   32: em_htg_other:*
   33: em_htg_mus:*
   34: em_htg_pln:*
   35: em_htg_rod:*
   36: em_htg_mam:*
   37: em_htg_vrt:*
   38: em_sy:*
   39: em_htgo_hum:*
   40: em_htgo_mus:*
   41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

---

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	555	100.0	555	6	AX399631	AX399631 Sequence
2	555	100.0	555	6	BD188111	BD188111 Method an
3	555	100.0	829	6	AX399629	AX399629 Sequence
4	555	100.0	552	6	BD188110	BD188110 Method an
5	550.4	99.2	2637	9	BD240900	BD240900 Expressio
6	419.6	75.6	546	6	BC033715	BC033715 Homo sapi
7	418	75.3	549	6	AR193165	AR193165 Sequence
8	418	75.3	549	6	BD240882	BD240882 Expressio
9	418	75.3	552	6	AX100086	AX100086 Sequence
10	418	75.3	552	6	BD266741	BD266741 Proteins
11	418	75.3	552	6	AX395662	AX395662 Sequence
12	418	75.3	552	6	AX473835	AX473835 Sequence
13	418	75.3	552	6	BD081407	BD081407 Fused pro
14	418	75.3	563	6	AR431862	AR431862 Sequence
15	418	75.3	786	9	AF282883	AF282883 Homo sapi
16	418	75.3	900	6	AX370851	AX370851 Sequence
17	418	75.3	1388	6	AK130035	AK130035 Homo sapi
18	418	75.3	2364	9	AK098216	AK098216 Homo sapi
19	418	75.3	2376	6	AX816106	AX816106 Sequence
20	418	75.3	2376	6	AX816108	AX816108 Sequence
21	418	75.3	3394	6	I51045	I51045 Sequence 4
22	418	75.3	3394	6	AX409531	AX409531 Sequence
23	418	75.3	3394	9	HUMCOL18AX	L22548 Human colla
24	418	75.3	5408	9	AF018082	AF018082 Homo sapi
25	418	75.3	5929	9	AF018081	AF018081 Homo sapi
26	416.4	75.0	564	9	AF416592	AF416592 Homo sapi
27	411.6	74.2	555	9	AF184060	AF184060 Homo sapi
28	410.8	74.0	534	6	AR193166	AR193166 Sequence
29	410.8	74.0	537	6	AX100092	AX100092 Sequence
30	398.4	71.8	650	6	AX370853	AX370853 Sequence
31	394.2	71.0	552	6	BD272245	BD272245 Anti-angi
32	379	68.3	851	10	AF189709	AF189709 Rattus no
33	375.8	67.7	555	6	BD102800	BD102800 Process f
34	375.8	67.7	555	6	BD128313	BD128313 Endostati
35	375.8	67.7	558	6	AR268609	AR268609 Sequence
36	375.8	67.7	565	6	AR100642	AR100642 Sequence
37	375.8	67.7	565	6	E34073	E34073 Carrier/DNA
38	375.8	67.7	624	6	AX084406	AX084406 Sequence
39	375.8	67.7	1615	10	MMU03714	U03714 Mus musculu
40	375.8	67.7	4437	10	MUSCOLLAG	D17546 Mus musculu
41	375.8	67.7	4802	10	BC043697	BC043697 Mus muscu
42	374.2	67.4	624	6	AX128409	AX128409 Sequence
43	374.2	67.4	4031	6	I51044	I51044 Sequence 1
44	374.2	67.4	4031	10	MUSCOLPRO	L22545 Mouse alpha
45	373.4	67.3	552	10	AF257775	AF257775 Mus muscu

ALIGNMENTS

RESULT 1
AX399631
LOCUS   AX399631   555 bp   DNA   linear   PAT 06-JUN-2002
DEFINITION Sequence 3 from Patent EP1191036.
ACCESSION   AX399631
VERSION   AX399631.1 GI:21335410
KEYWORDS
SOURCE   Canis familiaris (dog)
  ORGANISM Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1
  AUTHORS   Sheppard,M.G. and Tong,S.
  TITLE   Methods and compositions for diagnosing and treating disorders
            involving angiogenesis

Left column

```
JOURNAL      Patent: EP 1191036-A 3 27-MAR-2002;
             Pfizer Products Inc. (US)
FEATURES             Location/Qualifiers
     source          1..555
                     /organism="Canis familiaris"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9615"
ORIGIN

  Query Match             100.0%;  Score 555;  DB 6;  Length 555;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-64;
  Matches    555;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCCAGCCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCCAGCCG 60

Qy     61 GGCGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGGCCGCGCCGCCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGCGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGGCCGCGCCGCCG 120

Qy    121 GGGCTGGCCGGCACCTTCCGGGCCTTCGTGTCGCGGCTGCAGGACCTCTACAGCATC 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGGCTGGCCGGCACCTTCCGGGCCTTCGTGTCGCGGCTGCAGGACCTCTACAGCATC 180

Qy    181 GTGCGCGCGCCGACCGACCGCACCGGGGTGCCCGTGCGTCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTGCGCGCGCCGACCGACCGCACCGGGGTGCCCGTGCGTCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 240

Qy    241 CCCAGCTGGGAGGCCTTATTCTCGGGCTTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCCAGCTGGGAGGCCTTATTCTCGGGCTTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC 300

Qy    301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGGAGAGACGGTGTGG 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGGAGAGACGGTGTGG 360

Qy    361 CACGGCTCCGACCCCAGGGGCCACCGGGTCGTCCGCCTGGCGAGACGTGCGGACG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CACGGCTCCGACCCCAGGGGCCACCGGGTCGTCCGCCTGGCGAGACGTGCGGACG 420

Qy    421 GAGGCCCCGCGAGCTGCCGCCACCGGGCCAGGCGTCGTGCTGCTGCTCGGAGCAG 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GAGGCCCCGCGAGCTGCCGCCACCGGGCCAGGCGTCGTGCTGCTGCTCGGAGCAG 480

Qy    481 GAGGCCGGAGCTGCCGCCACCGGGCCACCGGTCCTGGTGCTCTGCATCGAGAACAGCGTCATGACC 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GAGGCCGGAGCTGCCGCCACCGGGCCACCGGTCCTGGTGCTCTGCATCGAGAACAGCGTCATGACC 540

Qy    541 TCCTTCTCCAAGTAG 555
          |||||||||||||||
Db    541 TCCTTCTCCAAGTAG 555


RESULT 2
BD188111
LOCUS       BD188111                555 bp    DNA    linear    PAT 17-JUL-2003
DEFINITION  Method and composition for diagnosing and treating diseases
            relating to angiogenesis.
ACCESSION   BD188111
VERSION     BD188111.1  GI:32997850
KEYWORDS    JP 2003000268-A/2.
SOURCE      Canis sp.
  ORGANISM  Canis sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Sheppard,M.G. and Tong,X.
  TITLE     Method and composition for diagnosing and treating diseases
            relating to angiogenesis
  JOURNAL   Patent: JP 2003000268-A 2 07-JAN-2003;
            PFIZER PRODUCTS INC
  OS        Canis sp. (dog)
COMMENT
```

Right column

```
PN   JP 2003000268-A/2
PD   07-JAN-2003
PF   24-AUG-2001 JP 2001254697
PR   25-AUG-2000 US    60/227924
PI   MICHAEL GEORGE SHEPPARD,XIAO TONG
PC   C12N15/09,A01K67/027,A61K31/7088,A61K35/76,A61K39/395 PC
     ,A61K39/395,A61K45/00,
PC   A61K48/00,A61P7/00,A61P9/00,A61P9/10,A61P15/00,A61P17/00, PC
A61P17/02,
PC   A61P17/06,A61P19/02,A61P19/08,A61P25/00,A61P27/02,A61P27/06,
PC   A61P29/00,
PC   A61P31/00,A61P35/00,A61P35/02,A61P35/04,A61P37/06,A61P43/00,
PC   C07K14/47,
PC   C07K16/18,C07K19/00,C12N1/15,C12N1/19,C12N1/21,C12N5/10 PC
     ,C12Q1/02,G01N33/15,
PC   G01N33/50,C12N15/00,C12N5/00
CC   CANINE ENDOSTATIN NUCLEOTIDE SEQUENCE
FH   Key             Location/Qualifiers
FT   source          1..555
FT                   /organism='Canis sp. (dog)'.

FEATURES             Location/Qualifiers
     source          1..555
                     /organism="Canis sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9616"
ORIGIN

  Query Match             100.0%;  Score 555;  DB 6;  Length 555;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-64;
  Matches    555;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCCAGCCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCCAGCCG 60

Qy     61 GGCGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGGCCGCGCCGCCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGCGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGGCCGCGCCGCCG 120

Qy    121 GGGCTGGCCGGCACCTTCCGGGCCTTCGTGTCGCGGCTGCAGGACCTCTACAGCATC 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGGCTGGCCGGCACCTTCCGGGCCTTCGTGTCGCGGCTGCAGGACCTCTACAGCATC 180

Qy    181 GTGCGCGCGCCGACCGACCGCACCGGGGTGCCCGTGCGTCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTGCGCGCGCCGACCGACCGCACCGGGGTGCCCGTGCGTCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 240

Qy    241 CCCAGCTGGGAGGCCTTATTCTCGGGCTTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCCAGCTGGGAGGCCTTATTCTCGGGCTTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC 300

Qy    301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGGAGAGACGGTGTGG 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGGAGAGACGGTGTGG 360

Qy    361 CACGGCTCCGACCCCAGGCGGCCACCGGGTCGTCCGCCTGCGCTGGGAGCAG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CACGGCTCCGACCCCAGGCGGCCACCGGGTCGTCCGCCTGCGCTGGGAGCAG 420

Qy    421 GAGGCCCCGCGAGCTGCCGCCACCGGGCCAGGCGTCGTGCTGCTGCTCGGAGCAG 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GAGGCCCCGCGAGCTGCCGCCACCGGGCCAGGCGTCGTGCTGCTGCTCGGAGCAG 480

Qy    481 GAGGCCGGAGCTGCGCCACGCGTCTGGTGGTGCTCTGCATCGAGAACAGCGTCATGACC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GAGGCCGGAGCTGCGCCACGCGTCTGGTGGTGCTCTGCATCGAGAACAGCGTCATGACC 540

Qy    541 TCCTTCTCCAAGTAG 555
          |||||||||||||||
Db    541 TCCTTCTCCAAGTAG 555

RESULT 3
```

```
AX399629
LOCUS       AX399629                 829 bp    DNA     linear   PAT 05-JUN-2002
DEFINITION  Sequence 1 from Patent EP1191036.
ACCESSION   AX399629
VERSION     AX399629.1  GI:21135409
KEYWORDS    
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1
  AUTHORS   Sheppard,M.G. and Tong,X.
  TITLE     Methods and composition for diagnosing and treating disorders
            involving angiogenesis
  JOURNAL   Patent: EP 1191036-A 1 27-MAR-2002;
            Pfizer Products Inc. (US)
FEATURES             Location/Qualifiers
     source          1..829
                     /organism="Canis familiaris"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9615"
ORIGIN

Query Match             100.0%;  Score 555;  DB 6;  Length 829;
Best Local Similarity   100.0%;  Pred. No. 1.1e-64;
Matches  555;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCCCTGAACAGCCCGCAGCCG   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   139 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCCCTGAACAGCCCGCAGCCG  198

Qy    61 GGGGGCATGGAGGCATCCGGGGAGCGGGACTTCCAGTTCTTCCAGCAGGCGGCGCCCGCG  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   199 GGGGGCATGGAGGCATCCGGGGAGCGGGACTTCCAGTTCTTCCAGCAGGCGGCGCCCGCG  258

Qy   121 GGGCTGGCCGGCCACCTTCCGGGCCTTCCTGTCGTCGCGGCGCTGCAGGACCTCTACAGCATC  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   259 GGGCTGGCCGGCCACCTTCCGGGCCTTCCTGTCGTCGCGGCGCTGCAGGACCTCTACAGCATC  318

Qy   181 GTGCGCGCCGCCGACCGCACCGGGGTGCCCGTGTCGTCAACCTCAGGGACGAGGTGCTCTTC  240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   319 GTGCGCGCCGCCGACCGCACCGGGGTGCCCGTGTCGTCAACCTCAGGGACGAGGTGCTCTTC  378

Qy   241 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   379 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC  438

Qy   301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCCTGGCCCCGGAAGAGCGTGTGG  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   439 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCCTGGCCCCGGAAGAGCGTGTGG  498

Qy   361 CACGGCTCCGACCCCCAGCGGGGCGCGCCCGGGGCGCCCAGCGGGGCGCGCCCGGGGCGC  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   499 CACGGCTCCGACCCCCAGCGGGGCGCGCCCGGGGCGCCCAGCGGGGCGCGCCCGGGGCGC  558

Qy   421 GAGGCCCGGCGGAGCTGCGCCACCGGGGCACCGGGGCGCGCCAGGGGCAGGCTGCTGAGCAG  480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   559 GAGGCCCGGCGGAGCTGCGCCACCGGGGCACCGGGGCGCGCCAGGGGCAGGCTGCTGAGCAG  618

Qy   481 GAGGCCGGAGCTGCCGCCACGCCCTTCGTGGTGCTCTGCAGAACAGGGCTCATGACC  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   619 GAGGCCGGAGCTGCCGCCACGCCCTTCGTGGTGCTCTGCAGAACAGCGTCATGACC  678

Qy   541 TCCTTCTCCAAGTAG  555
           |||||||||||||||
Db   679 TCCTTCTCCAAGTAG  693

RESULT 4
BD188110
LOCUS       BD188110                 829 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Method and composition for diagnosing and treating diseases
            relating to angiogenesis.
```

```
ACCESSION   BD188110
VERSION     BD188110.1  GI:32997849
KEYWORDS    JP 2003000268-A/1.
SOURCE      Canis sp.
  ORGANISM  Canis sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 829)
  AUTHORS   Sheppard,M.G. and Tong,X.
  TITLE     Method and composition for diagnosing and treating diseases
            relating to angiogenesis
  JOURNAL   Patent: JP 2003000268-A 1 07-JAN-2003;
            PFIZER PRODUCTS INC
COMMENT     OS   Canis sp. (dog)
            PN   JP 2003000268-A/1
            PD   07-JAN-2003
            PF   24-AUG-2001 JP 2001254697
            PR   25-AUG-2000 US    60/227924
            PI   MICHAEL GEORGE SHEPPARD,XIAO TONG
            PC   C12N15/09,A01K67/027,A61K31/7088,A61K35/76,A61K39/395 PC
            ,A61K39/395,A61K45/00,
            PC   A61K48/00,A61P7/00,A61P9/00,A61P9/10,A61P15/00,A61P17/00, PC
            A61P17/02,
            PC   A61P17/06,A61P19/02,A61P19/08,A61P25/00,A61P27/02,A61P27/06,
            PC   A61P29/00,
            PC   A61P31/00,A61P35/00,A61P35/02,A61P35/04,A61P37/06,A61P43/00,
            PC   C07K14/47,
            PC   C07K16/18,C07K19/00,C12N1/15,C12N1/19,C12N1/21,C12N5/10 PC
            ,C12Q1/02,G01N33/15,
            PC   G01N33/50,C12N15/00,C12N5/00
            CC   CANINE PRO ENDOSTATIN NUCLEOTIDE SEQUENCE
            FH   Key             Location/Qualifiers
            FT   source          1..829
            FT                   /organism='Canis sp. (dog)'.
FEATURES             Location/Qualifiers
     source          1..829
                     /organism="Canis sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9616"
ORIGIN

Query Match             100.0%;  Score 555;  DB 6;  Length 829;
Best Local Similarity   100.0%;  Pred. No. 1.1e-64;
Matches  555;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTCGCCCTGAACAGCCCGCAGCCG   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   139 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTCGCCCTGAACAGCCCGCAGCCG  198

Qy    61 GGGGGCATGGAGGCATCCGGGGAGCGGGACTTCCAGTGCTTCCAGCAGGCGGCGCCCGCG  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   199 GGGGGCATGGAGGCATCCGGGGAGCGGGACTTCCAGTGCTTCCAGCAGGCGGCGCCCGCG  258

Qy   121 GGGCTGGCCGCCACCTTCCGGGCCTGCTGCGCGTGCGCGTCGCGGCGCTGCAGGACCTCTACAGCATC  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   259 GGGCTGGCCGCCACCTTCCGGGCCTGCTGCGCGTGCGCGTCGCGGCGCTGCAGGACCTCTACAGCATC  318

Qy   181 GTGCGCGCCGCCGACCGCACCGGGGTGCCCGTGTCGTCAACCTCAGGGACGAGGTGCTCTTC  240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   319 GTGCGCGCCGCCGACCGCACCGGGGTGCCCGTGTCGTCAACCTCAGGGACGAGGTGCTCTTC  378

Qy   241 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAACAGCCCGCAGCCCGCATC  300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   379 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAACAGCCCGCAGCCCGCATC  438

Qy   301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCCTGGCCCCGGAAGAGCGTGTGG  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   439 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCCTGGCCCCGGAAGAGCGTGTGG  498

Qy   361 CACGGCTCCGACCCCCAGCGGGGCGCGCCCGACCGACAGCTACTGCGAGACGCGTGGCGGACG  420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   499 CACGGCTCCGACCCCCAGCGGGGCGCGCCCGACCGACAGCTACTGCGAGACGCGTGGCGGACG  558
```

```
Qy  421 GAGGCCCCGCGGCGGCCACCGGCGAGGCGTTCGTCGCTGCTGGCGGGGGCAGGCTCGCTGGAGCAG 480
        ||||||||  ||  ||||||||| |||||||| || || |||||||||| |||||||| ||||||||
Db  559 GAGGCCCCGGGCGGGGCCACCGGGCAGGCGTCGTGCTCGCTGCTGGCGGCGGCAGGCTGCTGGAGCAG 618

Qy  481 GAGGCCGGAGCTGCCGAGCTGCGCACGCTTCGTCGTTCGTGCTCTGCATCGAGAACAGCGTCATGACC 540
        ||||||||  |||||||  |||||||||| ||||||| |||||||||||||||||||||||||||||||
Db  619 GAGGCCGCGAGCTGCGCCACGCCTTCGTCGTGCTCTGCATCGAGAACAGCGTCATGACC 678

Qy  541 TCCTTCTCCAAGTAG 555
        |||||||||||||||
Db  679 TCCTTCTCCAAGTAG 693
```

```
RESULT 5
BD240900
LOCUS       BD240900    552 bp    DNA    linear    PAT 17-JUL-2003
DEFINITION  Expression and export of angiogenesis inhibitors as immunofusins.
ACCESSION   BD240900
VERSION     BD240900.1  GI:33050670
KEYWORDS    JP 2002523036-A/20.
SOURCE      Canis familiaris (dog)
ORGANISM    Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 552)
AUTHORS     Lo.K.M., Li Y. and Gillies,S.D.
TITLE       Expression and export of angiogenesis inhibitors as immunofusins
JOURNAL     Patent: JP 2002523036-A 20 30-JUL-2002;
            LEXIGEN PHARMACEUTICALS CORP
COMMENT     OS  Canis familiaris (dog)
            PN  JP 2002523036-A/20
            PD  30-JUL-2002
            PF  25-AUG-1999 JP 2000566305
            PR  25-AUG-1998 US 60/097883
            PI  KIN MING LO,YUE LI,STEPHEN D GILLIES
            PC  C12N15/09,A61K31/711,A61K38/00,A61K39/395,A61K48/00,A61P7/00,
            PC  A61P35/00,C07K14/745,C07K14/78,C07K19/00,C12N5/10,C12P21/02,
            PC  C12N15/00,
            PC  C12N5/00,A61K37/02
            CC  Endostatin
            FH  Key             Location/Qualifiers
            FT  CDS             (1)..(552).
FEATURES             Location/Qualifiers
     source          1..552
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9615"
ORIGIN

Query Match           99.2%;  Score 550.4;  DB 6;  Length 552;
Best Local Similarity 99.8%;  Pred. No. 5e-64;
Matches 551;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;
```

```
Qy    1 CACACCCACCAGGACTTCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 60
        |||||| |||||||||||| ||| || |||||||||||||||||||||||||||||||
Db    1 CACACCCACCAGGACTTCCAGCGCTGCTGCACCCTGGTGCCCTGAACAGCCCGCAGCCG 60

Qy   61 GGGCGGCATGCGGAGGCATCCGGGAGGCCGACTTCCAGTGCTTCCAGGAGGCGCGCGCCGCGG 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGGCGGCATGCGGAGGCATCCGGGAGGCCGACTTCCAGTGCTTCCAGGAGGCGCGCGCCGCG 120

Qy  121 GGGCTGGCCGGCGACCTTCCGGGCCTTCCGTCGTCGTGCGGGTGCAGGACCTCTACAGCATC 180
        |||||||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||
Db  121 GGGCTGGCCGGCGACCTTCCGGGCCTTCCGTCGTCGCGGGTGCAGGACCTTCTACAGCATC 180

Qy  181 GTGCGCCGGCGCCGACCGGCGCACCCGGGGTGCCCGTCGTCAACCTCAGGGACGAGGGTGCTCTTC 240
        |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db  181 GTGCGCCGGCGCCGACCGGCGCACCCGGGGGTGCCCGTCGTCAACCTCAGGGACGAGGGTGCTCTTC 240

Qy  241 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGGCCCGGGGGCCCGCATC 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGGCCCGGGGGCCCGCATC 300
```

```
Qy  301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCGGCTGGCCCGGAAAGAGCGTGTGG 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCGGCTGGCCCGGAAAGAGCGTGTGG 360

Qy  361 CACGGCTCCGACCCCAGCGGCGGCCGCCCTGACCGACACGCACTGCGAGACGTCGCGGACG 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 CACGGCTCCGACCCCAGCGGCGGCCGCCCTGACCGACACGCACTGCGAGACGTCGCGGACG 420

Qy  421 GAGGCCCCGGCGGCCACCGGGCAGGCAGGCGTGTCGCTGCTGGCGGGCAGGCTGCTGGACCAG 480
        ||||||||||||||||||||||||||| |||||| ||||||||||||||||||||||||| |||
Db  421 GAGGCCCCGGCGGCCACCGGGCAGGCAGGCGTCGTGCTCGCTGCTGGCGGGCAGGCTGCTGGAGCAG 480

Qy  481 GAGGCCGGCGAGCTGCCGCCAGCTTCGTCGTGCTCTGCATCGAGAACAGCGTCATGACC 540
        ||||||||||||||||  ||  ||||||||||||||||||||||||||||||||||||||
Db  481 GAGGCCGGCGAGCTGCGCCACGCCTTCGTCGTGCTCTGCATCGAGAACAGCGTCATGACC 540

Qy  541 TCCTTCTCCAAG 552
        ||||||||||||
Db  541 TCCTTCTCCAAG 552
```

```
RESULT 6
BC033715
LOCUS       BC033715    2637 bp    mRNA    linear    PRI 08-JUL-2002
DEFINITION  Homo sapiens, similar to collagen, type XVIII, alpha 1, clone
            IMAGE:4425380, mRNA, partial cds.
ACCESSION   BC033715
VERSION     BC033715.1  GI:21708148
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2637)
AUTHORS     Strausberg,R.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUL-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Makkeri,B., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 68 Row: e Column: 16
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.
FEATURES             Location/Qualifiers
     source          1..2637
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4425380"
```

Query Match   75.3%; Score 418; DB 6; Length 546;
Best Local Similarity 85.3%; Pred. No. 1.9e-46;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy     1 CACACCCACCAGGACTTCCAGTGTGTCGTGCACCTGGTGCCCTGAACAGCCCGCAGCCG  60
            |||||||||||||||||||||||||||||||||| |||||||||||||| ||||||||||
Db     1 CACAGCCACCACGCGGACTTCCAGCGTCCACCTGGTTGCCCTCAACAGCCCCCTCTCA  60

Qy    61 GGGGCCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGCGCGGCGCCGCGCG 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GGGGCCATGCGGGCGGGGCATCCGCGGGGCCGACTTCCAGTGCTTCCAGCAGCGCGGCGGCGTG 120

Qy   121 GGGCTGGCCGGCGGACTTCCGGGCGCTTCCGTGCTCGTGCGGGCTGCAGCACTTCCAGCAGCATC 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGGCTGGCGGGGGCACCTTCCGCGCCTTCGTGCTGGGGCCTGCAGGACCTGCAGCAGCATC 180

Qy   181 GTGCGCCGCGCCGACCGGTCCGTGCGGGCTCAACTTCAGGGAGGCAGGGTGTCTTC 240
            |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTGCGCCGTGCGGACCGCGCCGAGCCGTGCCATCGTCAACTTCAAGGACGAGCAGCTGTGTTT 240

Qy   241 CCCAGCTGGAGGCCTTATTCTCCGGGCTCGGAGAGGCTGAAGCCCGGGCCCGATC 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CCCAGCTGGGAGGCTCGTCGTTCTCAGGCTCGAGGGTCCGTCGAGGGCACGCATC 300

Qy   301 TTCTCTTTCGACGGCAGAGATGTCCTGGCACGGGCCGGGGCCGACGAGCGGTGTGG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TTCTCCTTTGACGGCAAGGACGTTCCTGAGGCACGGGCTCCGAGCGAGCGGTGTGG 360

Qy   361 CACGGCTTCGACCCCCAGCCCAGCGGCAGGGTCCGTGCTGCTGGCCGGACG 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CATGGCTTCGACCCCCAACGGGCGGCAGGGCTACTGCGGCGCCCGAGCGACG 420

Qy   421 GAGGCCCCGGCAGCTGCCCGCCACCGCCTTCGTGGTGCTCGTCGGGGCAG 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAGGCTCCCTGGGCCACCGGGGCCAGGGGCCTCCTGCTCGTCTCGGGGCAG 480

Qy   481 GAGGCCGGCAGCTGCCCGCCACCGCCTTCCGTGGTGCTCCGTGGTCTGATGACC 540
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 AGTGCGCGGCACAGCTGCCCATCACGCCTACATCGTGCTCGCCTTCATGACT 540

Qy   541 TCCTTC 546
            ||||||
Db   541 GCCTCC 546
```

CDS (left column)

```
/tissue_type="Kidney, hypernephroma"
/clone_lib="NIH_MGC_89"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
<1..2453
/codon_start=3
/product="similar to collagen, type XVIII, alpha 1"
/protein_id="AAH33715.1"
/db_xref="GI:21708149"
/translation="TRPPGREGPPGPFGPGLPGPPGPGREGPPGPPGPGLPAGFDDMEGSGGPFW
GHKGSKGAPGPAGARGESGLAGAPGPAGPPGPPGPGPGPGLPAGFDDMEGSGGPFW
STARSADGPQGPPGLPGLKGDPGVPGLPGAKGEVGADGIPGFPGLPGREGIAGPQGPK
GDRGSRGEKGDPGKDGVQPGLPGPGPGPGPVYSEQDGSVLSVPGPGRPGFAGFP
GPAGPKGNLGSKGERGSPGPKGEKGEPGGIFSPDGGALGPAQKGAKGEPGFRGPPGPY
GRPGYKGEIGFPGRPGRPGMNGLKGEKGEPGDASLGFGMKGMFGPPGPPGPPGTP
VYDSNVFAESSRPGPPGLPGNQGFPGPGKGAKGEVGPPGPPGQFPDFLQLEAEMKGEK
GDRGDAGQKGERGEBPGGGGFFGSSLPGPPGPRGYPGIPGPKGESIRGQPGPPGPQ
GPPGIGYEGRGQPGPPGPPGPPSFPGPHRQTISVPGPPGPGTMGASSGVR
LWATRQAMLGQVHEVPGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRGTNEVAA
LQPPVVQLHDSNPYPRREHPHPTARPWRADDILASPPRLPEPOPYPGAPHHSSYVHLR
PARPTSPPAHSHRDFPQPVLHIVALNSPLSGGMRGIRGADFDQCFQQARAVGLAGTFRAF
LSSRLQDLYSIVRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKGPGKGARIFSFDGKD
VLRHPTWPQKSVWHGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASC
HHAYIVLCIENSFMPASK"
ORIGIN
```

Query Match   75.6%; Score 419.6; DB 9; Length 2637;
Best Local Similarity 85.5%; Pred. No. 7e-47;
Matches 467; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

```
Qy     1 CACACCCACCAGGACTTCCAGTGTGTCGTGCACCTGGTGCCCTGAACAGCCCGCAGCCG    60
            |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db  1902 CACAGCCACCACGCGGACTTCCAGCGTCCACCTGGTTGCCCTCAACAGCCCCCTGTCG  1961

Qy    61 GGGGCCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGCGCGGCGCCGCG   120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1962 GGGGCCGGCATGCGGGCGGGGCATCCGCGGGGCCGACTTCCAGTGCTTCCAGCAGCGCGGCGGCCGTG  2021

Qy   121 GGGCTGGCCGGCGCACCTTCCGGGCGGCCTTCCGTGCTCGTGCGGGCTGCAGGACCTCTACAGCATC  180
            ||||||||||||||||||||||||||||||||||||||||||||||
Db  2022 GGGCTGGCGGGGGCACCTTCCGCGCCTTCCGTGCTCCGCGCGCCTGCAGGACCTGTACAGCATC  2081

Qy   181 GTGCGCCGCGCCGACCGGTCCGTGCGGGCTCAACTTCAGGGAGGCAGGGTCCTACAGCATC  240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2082 GTGCGCCGTGCGGACCGCGCCGAGCCGTGCCATCGTCAACTTCAAGGACGAGCAGCTGTGTTT  2141

Qy   241 CCCAGCTGGGAGGCCTTATTCTCCGGGCTCGGAGAGGCTGAAGCCCGGGCCCGATC  300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2142 CCCAGCTGGGAGGCTCGTCGTTCTCAGGCTCGAGGGTCCGTCGAGGGCACGCATC  2201

Qy   301 TTCTCTTTCGACGGCAGAGATGTCCTGGGCACGGGCCGGGGCCGACGAGCGGTGTGG  360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2202 TTCTCCTTTGACGGCAAGGACGTCCTGAGGCACGGGCTCCGAGCGAGCGGTGTGG  2261

Qy   361 CACGGCTCCGACCCCCAGCCGCGGCCGGGGCCTGACCGACCAGCTACTGCGGGACG  420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2262 CATGGCTCCGACCCCCAACGGGCCCAGCGGCCGCCAGCGAGAGCTACTGGCCGGACG  2321

Qy   421 GAGGCCCCGGCAGCTGCCCGCCACCGCCTTCGTGGTGCTCGTCGGGGCAG  480
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2322 GAGGCTCCCTGCGGCCACCGGGGCCAGGGGCCTCCTGCTGGGGGCAG  2381

Qy   481 GAGGCCGGCAGCTGCGCGCCACCGCCTTCGTGGTGCTCGTCGGGGCAG  540
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  2382 AGTGCGCGGCAGCTGCCATCACGCGCTACATCGTGCTCTGCATTGAGAACAGCGTTCATGACT  2441

Qy   541 TCCTTC 546
            ||||||
Db  2442 GCCTCC 2447
```

RESULT 7
AR193165

Left column

```
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Lo,K.M., Li,Y. and Gillies,S.D.
  TITLE     Expression and export of angiogenesis inhibitors as immunofusins
  JOURNAL   Patent: JP 2002523036-A 2 30-JUL-2002;
            LEXIGEN PHARMACEUTICALS CORP
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002523036-A/2
            PD   30-JUL-2002
            PF   25-AUG-1999 JP 2000556305
            PR   25-AUG-1998 US  60/097883
            PI   KIN MING LO,YUE LI,STEPHEN D GILLIES
            PC   C12N15/09,A61K31/711,A61K38/00,A61K39/395,A61K48/00,A61P7/00,
            PC   A61P35/00,
            PC   A61P43/00,C07K14/745,C07K14/78,C07K19/00,C12N5/10,C12P21/02,
            PC   C12N15/00,
            PC   C12N5/00,A61K37/02
            CC   endostatin
FH   Key              Location/Qualifiers
FT   CDS              (1)..(549).
FEATURES             Location/Qualifiers
     source          1..549
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match             75.3%;  Score 418;  DB 6;  Length 549;
Best Local Similarity   85.3%;  Pred. No. 1.9e-46;
Matches  466;  Conservative   0;  Mismatches   80;  Indels   0;  Gaps   0;

Qy    1 CACACCCACCAGGACTTCCAGCTGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG  60
          ||||| ||||| ||||||||||||  ||||||||||||| |||| ||||||||||
Db    1 CACAGCCACCACGGCGACTTCCAGCCGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA   60

Qy   61 GGGCGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCGTGCGGCGCGCCGCCGCG  120
          |||||||||||||||||||  || | |||||||||||||||       |||||
Db   61 GGGCGCATGCGAGGCATCCGCGGGGCCGACTTCCAGTGCTTCCAGCAGCGCGCGGCCGTG  120

Qy  121 GGGCTGGCCGGCCACCTTCCTGGGCGCTCGCGGCGCCTGCGAGGAGGACCTCTACAGCATC 180
          ||||||| ||||  ||||||||| | |||||||||||||||||||  |||| ||||||||
Db  121 GGGCTGGCGGGCGGGCCACCTTCCTGTCTCGCGGCGCCTGCGAGGACCTGTACAGCATC   180

Qy  181 GTGCGGCGCGCCGACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGAGTGCTCTTC       240
          |||||| |  |||||||| | |||  ||||||||||| |||||||||| |||| ||
Db  181 GTGCGGCCGTGCCGACCGCGGCCCGCCATCGTCAACCTCAAGGACGAGAGCTGCTGTTT   240

Qy  241 CCCAGCTGGGAGGCCCTTATTCTCCGGGGCTCGAGCCCGGGCCCCGGAAGCCCGAGCCATC 300
          ||||||||||||| | | | |||  |||| |||||||| ||||  ||||| |||||||||
Db  241 CCCAGCTGGGAGGCTCGTTCTCAGGCTCTGAGGGACGAGAGCCCGGCCCACGCGCACGCATC 300

Qy  301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGGCTTGCGAGACGTGGCGGACG     360
          |||||  || ||||||| |||  |||||  | || || | |||||| ||||||||||
Db  301 TTCTCCTTTGACGGCAAGAGCGTCCTGAGCGACGTCGAGAGCGTGGCGGACG            360

Qy  361 CACGGCTCCGACCCCGGCGGCCAGCCAGCGGCCACCGGCGCCTCCTGGGGCAG          420
          ||||||| ||||||  |||| |||||| || | |||| || |||||||||||
Db  361 CATGGGCTCGGACCCCCACGGGCGCAGCCAGCGGCCAGCGGGCTCCTGGGGCAG          420

Qy  421 GAGGCGCGAGCTGCCGCCCACGCCTTCGTGGTGCTCGCTGCTGCGGCGGGCAG          480
          ||||| ||||||| |||||||| || | |||  ||| |||||  ||| | |||
Db  421 GAGGCTCCTCGGCACGGGCCACCGGGGCTCCTGGGGCAGCTTCATGACTCATGGG       480

Qy  481 GAGGCGCGAGCTGCCGCCCACGCCTTCGTGGTGCTCGCTGCTGCGGCGGGCAG          540
Db  481 AGTGCCGCGAGCTGCCATCACGCCTACATCAGCGTCTCTGCGTTCATGACT            540

Qy  541 TCCTTC 546
Db  541 GCCTCC 546
```

RESULT 9

Right column

```
            AX100086
LOCUS       AX100086                 549 bp    DNA     linear   PAT 02-APR-2001
DEFINITION  Sequence 4 from Patent WO0119989.
ACCESSION   AX100086
VERSION     AX100086.1  GI:13539061
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Liang,H., Sim,K.L., Chang-Murad,A., Zhou,X., Madsen,J.,
            Boerner,R.J., Bermejo,L.L., Mistry,F.R., Shepard,S.R. and
            Schrimsher,J.L.
  TITLE     Method of producing and purifying endostatin?tm protein
  JOURNAL   Patent: WO 0119989-A 4 22-MAR-2001;
            EntreMed, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..549
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match             75.3%;  Score 418;  DB 6;  Length 549;
Best Local Similarity   85.3%;  Pred. No. 1.9e-46;
Matches  466;  Conservative   0;  Mismatches   80;  Indels   0;  Gaps   0;

Qy    1 CACACCCACCAGGACTTCCAGCTGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG  60
          ||||| ||||| |||||||||||||  ||||||||||||| |||| ||||||||||||
Db    1 CACAGCCACCACGGCGACTTCCAGCCGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA   60

Qy   61 GGGCGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCGTGCGGCGCGCCGCCGCG  120
          |||||||||||||||||||  || | |||||||||||||||        |||||
Db   61 GGGCGCATGCGAGGCATCCGCGGGGCCGACTTCCAGTGCTTCCAGCAGCGCGCGGCCGTG  120

Qy  121 GGGCTGGCCGGCCACCTTCCTGGGCGCTCGCGGCGCCTGCGAGGAGGACCTCTACAGCATC 180
          ||||||| ||||  ||||||||| | |||||||||||||||||||  |||| ||||||||
Db  121 GGGCTGGCGGGCGGGCCACCTTCCTGTCTCGCGGCGCCTGCGAGGACCTGTACAGCATC   180

Qy  181 GTGCGGCGCGCCGACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGAGTGCTCTTC       240
          |||||| |  |||||||| | |||  ||||||||||| |||||||||| |||| ||
Db  181 GTGCGGCCGTGCCGACCGCGGCCCGCCATCGTCAACCTCAAGGACGAGAGCTGCTGTTT   240

Qy  241 CCCAGCTGGGAGGCCCTTATTCTCCGGGGCTCGAGCCCGGGCCCCGGAAGCCCGAGCCATC 300
          ||||||||||||| | | | |||  |||| |||||||| ||||  ||||| |||||||||
Db  241 CCCAGCTGGGAGGCTCGTTCTCAGGCTCTGAGGGACGAGAGCCCGGCCCACGCGCACGCATC 300

Qy  301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGGCTTGCGAGACGTGGCGGACGTGG  360
          |||||  || ||||||| |||  |||||  | || || | |||||| ||||||||||||||
Db  301 TTCTCCTTTGACGGCAAGAGCGTCCTGAGCGACGTCGAGAGCGTGGCGGACGTGG        360

Qy  361 CACGGCTCCGACCCCGGCGGCCAGCCAGCGGCCACCGGCGCCTGCTGGGGCGGACG       420
          ||||||| ||||||  |||| |||||| || | |||| || ||||||||||||||||
Db  361 CATGGGCTCGGACCCCCACGGGCGCAGCCAGCGGCCAGCGGGCTGCTGGGGCGGACG      420

Qy  421 GAGGCGCGGCCGCCACGCCTTCGTGGTGCTGCTGGGCGGACGTGGCGGACGCAG          480
          ||||| ||||| |||| || | ||| |||| | || |||||||| || || |||
Db  421 GAGGCTCCTCGGCACGGGCCACCGGGGCTCCTGGGGCAGCTTCATGACTCATGGG        480

Qy  481 GAGGCGCGAGCTGCCGCCCACGCCTTCGTGGTGCTCGCTGCTGCGGCGGGCAG          540
          |||||  |||||| |||||  || | |||| |||  |||||||||||||| |||
Db  481 AGTGCCGCGAGCTGCCATCACGCCTACATCAGCGTCTCTGCGTTCATGACT            540

Qy  541 TCCTTC 546
Db  541 GCCTCC 546
```

RESULT 10
BD266741
LOCUS   BD266741   552 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION Proteins that bind angiogenesis-inhibiting proteins, compositions

```
ACCESSION    BD266741
VERSION      BD266741.1  GI:33076509
KEYWORDS     JP 2002532068-A/2.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 552)
  AUTHORS    Macdonald,N.J. and Sim,K.L.
  TITLE      Proteins that bind angiogenesis-inhibiting proteins, compositions
             and methods of use thereof
  JOURNAL    Patent: JP 2002532068-A 2 02-OCT-2002;
             ENTREMED INC
COMMENT      OS   Homo sapiens (human)
             PN   JP 2002532068-A/2
             PD   02-OCT-2002
             PF   06-DEC-1999 JP 2000585271
             PR   04-DEC-1998 US   09/206059
             PI   NICHOLAS J MACDONALD,KIM LEE SIM
             PC   C12N15/09,A61K38/00,A61P9/00,A61P27/02,A61P29/00,A61P35/00, PC
             A61P43/00,
             PC
             C07K14/705,G01N33/15,G01N33/50,G01N33/566//A61K35/76, PC
             A61K45/00,
             PC   A61K48/00,C12N15/00,A61K37/02
             CC   Proteins that bind angiogenesis-inhibiting proteins,    CC
             compositions and
             CC   methods of use thereof
             FH   Key             Location/Qualifiers
             FT   source          1..552
             FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1..552
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match        75.3%;  Score 418;  DB 6;  Length 552;
Best Local Similarity   85.3%;  Pred. No. 1.9e-46;
Matches 466;  Conservative   0;  Mismatches  80;  Indels   0;  Gaps    0;

Qy       1 CACACCCACCAGGACTTCCAGTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 60
           |||||||  ||| ||||||||||| |||||||||||||||| ||| |||||||||||
Db       1 CACAGCCACCGCGACTTCCAGCCGGTCGTGCTCCACCTGGTTGCGCTCAACAGCCCCTGTCA 60

Qy      61 GGCGGCATGCGAGGCATCCGGGGAGCCGACTTCCAGTGCTTCGGCGGCGCGCGCGCCG 120
           |||||||||||||||| ||||||||||| |||||||||||||||||||||||||||||
Db      61 GGCGGCATGCGGGGGCATCGCGGGGCCGACTTCCAGTGCTTCGGCGGCGCGGGGCCGTG 120

Qy     121 GGCTGCGCCGGCCACCTTCCGGGGCCTTCCTGTCGTCGGGGTGCCCGTCGTCAACCTCACAGCATC 180
           ||||| ||||||||||||||||| ||| |||||||||||||||||||||||||||||
Db     121 GGCTGGCGCGTGCGGGGCCACCTTCCGCGGCCTTCCTGTCCTGCGGGTGCCCGTCGTCAACCTCGTACAGCATC 180

Qy     181 GTGCCGCGGCCACCGGACCGGGGCCTTGCCCGTCGTCGTCCGTGTCGGGGTGCCGTCGCCACCGGGTGTGG 240
           ||||||||||||||||||  ||||||| |||||||||||||||||||||||||||||
Db     181 GTGCCGCCGTGCGGACCGGGGCCTTGCCCCGTCGTCGGGGTGCCCGTCGTCGTGGTGTGG 240

Qy     241 CCCAGCTGGGAGGCCTTATTCTCGGGGCTCGAGATGTCCTGCAGAGCCCGGGCCCCATC 300
           ||||||||||||  ||||| ||||||||||||||||||||||||||||||||||||
Db     241 CCCAGCTGGGAGGCCTTATTCTCCGGGCTCCGAGATGTCCTGCAGAGCACCCCACCTGCCCCCAGAGCCCGGGCCCACGGCACCCATC 300

Qy     301 TTCTCTTTGACGGCAGAGATGTCCTGGCGTGG 360

Qy     361 CACGGCTCCGACCCCAGCCCGGCGCGCCCAGCCAGAGACTGTGGCGGACG 420

Qy     421 GAGGCGCCCCCGGCGCCCGGCCACCGGGCCAGGCGTGCCGCCACGCCTGCTGGTGTGG 480
```

RESULT 11
```
AX395662
LOCUS        AX395662                552 bp    DNA     linear   PAT 18-MAY-2002
DEFINITION   Sequence 30 from Patent WO0193897.
ACCESSION    AX395662
VERSION      AX395662.1  GI:21066471
KEYWORDS     .
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1
  AUTHORS    Sim,K.L. and Macdonald,N.J.
  TITLE      Angiostatin and endostatin binding proteins and methods of use
  JOURNAL    Patent: WO 0193897-A 30 13-DEC-2001;
             EntreMed, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..552
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match        75.3%;  Score 418;  DB 6;  Length 552;
Best Local Similarity   85.3%;  Pred. No. 1.9e-46;
Matches 466;  Conservative   0;  Mismatches  80;  Indels  80;  Gaps    0;

Qy       1 CACACCCACCAGGACTTCCAGTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 60
           |||||  |||| ||||||||||| |||||||||||||| |||| ||||||||||||
Db       1 CACAGCCACCGCGACTTCCAGCCGGTCGTGCTCCACCTGGTTGCGCTCAACAGCCCCTGTCA 60

Qy      61 GGCGGCATGCGAGGCATCCGGGGAGCCGACTTCCAGTGCTTCGGCGGCGCGCGCGCCG 120
           |||||||||||||||  ||||||||||| |||||||||||||||||||||||||||||
Db      61 GGCGGCATGCGGGGGCATCGCGGGGCCGACTTCCAGTGCTTCGGCGGCGCGGGGCCGTG 120

Qy     121 GGGCGGCGCGGGGACCTTCCTGTCGTCGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 180
           |||||| ||||||||  ||||||||||||||||||||||||| ||||||| |||||||
Db     121 GGGCTGGCGGGGGCACCTTCCGCGGCCTTCCTGTCCTGCGGGTGCCCGTCGTCGTCAACCTCAAGGACGAGCTGCTGTT 180

Qy     181 GTGGGCGGCGGCGACCGGACCGGGGCGGCCAGCCTGCCATCGTCAACCTCAAGGACAGAGCTGCTGTT 240
           ||||||||||||||||||||  ||||||||||||| ||||||||||||| |||||||
Db     181 GTGGGCCGTGCGGACCGGGGCCTCCAGGCTCTGAGGGTCCTCGAGGGGTCCGGAGGCCACCCGGTGTGG 240

Qy     241 CCCAGCTGGGAGGCCTTATTCTCGGGGCTCCGAGATGTCCTGCAGAGCCCGGGCCCCATC 300
           ||||||||||||  ||||| |||||||||||||||||||||||||||||||||||||
Db     241 CCCAGCTGGGAGGCTCAAGAACGTCTCCAGGCTCTGAGGCACCCCACCTGCCCCCAGAGCACCGGGCCCCATC 300

Qy     301 TTCTCTTTGACGGCAGAGATGTCCTGGCGTGG 360
           ||||||| |||||||||||||||||||||||
Db     301 TTCTCCTTTGACGGCGACCGGCAAGGACGTCTCCAGGCTCTGAGGCGGAGGCGGGCGTGG 360

Qy     361 CACGGCTCCGACCCCAGCCCGGCGCGCCCAGCCAGAGACTGTGGCGGACG 420
           |||||||||||||||||||||||||||||||||||||||||| |||||||
Db     361 CATGGCTCGGACCCCAGCCCGGCGCGCCCAGCCACTGGCGGCAGCTACTGTGATGACGTGGGCGGACG 420

Qy     481 GAGGCGCGGAGCTGCGCCACGCCTGCTGGGGGGCAGCCAGGCGCTCCTGGGGGCAG 540
           |||||  ||||||||||||||||| |||| ||||
Db     481 AGTGCCGCGGAGCTGCCATCACGCCTACATCGTCGTCCTGCATTGAGAACAGCTTCATGACT 540
```

```
Db      421 GAGGCTCCCTCCGGCCCAGGGGCCAGGCGCTCCTGCGTCTGTGGGGGCAG 480
Qy      481 GAGGCCGCGAGCTGCGCCACGCCTGCCATCACGCCTACATCGTCGTCCTGCATTGAGAACACAGCTTCATGACC 540
Db      481 AGTGCCGCGAGCTGCCATCACGCCTACATCGTCGTCCTGCATTGAGAACAGCTTCATGACT 540
Qy      541 TCCTTC 546
            ||||
Db      541 GCCTCC 546
```

Qy   541 TCCTTC 546
Db   541 GCCTCC 546

RESULT 12
AX473835
LOCUS   AX473835   552 bp   DNA   linear   PAT 09-AUG-2002
DEFINITION Sequence 2 from Patent WO0230982.
ACCESSION   AX473835
VERSION   AX473835.1 GI:22208005
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Sim,K.L. and Liang,H.
TITLE   Angiogenesis-inhibiting peptides and proteins and methods of use
JOURNAL   Patent: WO 0230982-A 2 18-APR-2002;
            EntreMed, Inc. (US)
FEATURES   Location/Qualifiers
source   1..552
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match   75.3%; Score 418; DB 6; Length 552;
Best Local Similarity 85.3%; Pred. No. 1.9e-46;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy   1 CACACCCACCAGGACTTCCAGCTGTGTCACCTGGTGCCCTGAACAGCCCGCAGCCG 60
           |||| |||| ||||||||||||| ||| |||||||||| ||||||||| |||
Db   1 CACAGCCACCGCGGACTTCCAGCCGGTGCTCCACCTGGTTGCGGCTCAACAGCCCCCTGTCA 60

Qy   61 GGCGGCATGCGAGGCATCCGGGGAGCCGACTTCCAGTGCTTCCAGCAGGCGCGGCCGCG 120
           ||||||||||||||||||||| |||||||||||||||||||||||||| |||||||||
Db   61 GGCGGCATGCGGGGCATCCGGGGCCGACTTCCAGTGCTTCCAGCAGGCGCGGCCGCGGTG 120

Qy   121 GGGCTGGCCGCGCACCTTCGGGGCTTCCTGTGTCGGCGGACCTCTACAGCATC 180
           ||||||| |||||||||||| ||||||||||||||||||||||||||||||||
Db   121 GGGCTGGCGGACCCCAACGGGCCGCAGGCTGCACTGTGAGACGTGGCGGACGGTGTACAGCATC 180

Qy   181 GTGCGCGCGCCGACCGACCCGGGTGCCCGTGTCAACCTCAAGGACGAGGTGCTCTTC 240
           |||||||||||||||||| ||||||||| |||||||||||||| ||||||||| |||
Db   181 GTGCGCGCGTGCCGACCGCGCAGCCGTGCTGCCGTCGTGTCAACCTCAAGGACGAGCTGCTGTTT 240

Qy   241 CCCAGCTGGGAGGGCCTTATTCTCCGGGCTCCGAGGGCCCAGCTGAAGCCCGGGGCCCGCATC 300
           |||||||||||| ||| | |||||||||| ||||||||||||| |||||||||||||
Db   241 CCCAGCTGGGAGGGCTCTGTTCTCAGGGTCCGGTTGAAGCCCGGGGGCACGCATC 300

Qy   301 TTCTCTTTGACGGCAGAGATGTCTCGAGCACCCCGGCCCTGCCCCTGTGG 360
           ||||| ||||| ||||| | |||||||||||||||| ||||||||||||
Db   301 TTCTCTTTGACGGCAGGAAGACGTCTCTGAGGCACCCCAGACCGTGGCCTGTGG 360

Qy   361 GGGCTCCGACCCCAGCCGGCGCGGCCGCCGGGCCGACCTCGGCCGACGTGGCGGACG 420
           || |||| ||||||||| || ||||||||||| |||||||| ||||||||||
Db   361 CATGGCTCCGGACCCCAACGGGCCGCAGGCTACTGTGAGACGTGGCGGACG 420

Qy   421 GAGGCCCCGGCGGGGCCACCGGGCAGGGCGCTCGTCGTGCTGGGGGCAG 480
           || ||||||| |||| |||||||| |||||||| || ||||||||
Db   421 GAGGCTCCCTGCGCCCAAGGGCCGCGGCCTCCTCGCTGCTCGGGGGCCAG 480

Qy   481 GAGGCCGGCAGCTGCCGCCAACGCGCCTTCGTGGTGCTCTGCATCGAGAGACAGCGTCATGACC 540
           |||||| ||||||||||||||| || || |||||||||||||||||||||||| ||||||
Db   481 AGTGCCGCAGCTGCCATCACGCCTACATCGTGTCGCATTGAGAACAGCTTCATGACT 540

Qy   541 TCCTTC 546
           || ||
Db   541 GCCTCC 546

RESULT 13
BD081407
LOCUS   BD081407   552 bp   DNA   linear   PAT 27-AUG-2002
DEFINITION Fused protein containing angiostatin component and utilization
            thereof in antitumor therapy.
ACCESSION   BD081407
VERSION   BD081407.1 GI:22627010
KEYWORDS   JP 2001518304-A/50.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 552)
AUTHORS   Bolanowski,M.A., Caparon,M.H., Casperson,G.F., Gregory,S.A.,
            Klein,B.K. and Mckearn,J.P.
TITLE   Fused protein containing angiostatin component and utilization
            thereof in antitumor therapy
JOURNAL   Patent: JP 2001518304-A 50 16-OCT-2001;
            GD SEARLE AND CO
COMMENT   OS   Homo sapiens (human)
            PN   JP 2001518304-A/50
            PD   16-OCT-2001
            PF   30-SEP-1998 JP 2000513958
            PR   01-OCT-1997 US 60/060609
            PI   MARK A BOLANOWSKI,MAIRE H CAPARON,GERALD P CASPERSON,SUSAN A
            PI   GREGORY,
            PI   BARBARA K KLEIN,JOHN P MCKEARN
            PC   C12N15/09,A61K38/00,A61K48/00,A61P9/10,A61P35/00,C07K14/52, PC
            C07K14/56,
            PC   C07K14/78,C12N9/68,C12N15/00,A61K37/02
            CC   Fused protein containing angiostatin component and utilization
            CC   thereof in
            CC   antitumor therapy
FH   Key   Location/Qualifiers
FT   source   1..552
                            /organism='Homo sapiens (human)'
FEATURES   Location/Qualifiers
source   1..552
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match   75.3%; Score 418; DB 6; Length 552;
Best Local Similarity 85.3%; Pred. No. 1.9e-46;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy   1 CACACCCACCAGGACTTCCAGCTGTGTCACCTGGTGCCCTGAACAGCCCGCAGCCG 60
           |||| |||| ||||||||||||| ||| |||||||||| ||||||||| |||
Db   1 CACAGCCACCGCGGACTTCCAGCCGGTGCTCCACCTGGTTGCGGCTCAACAGCCCCCTGTCA 60

Qy   61 GGCGGCATGCGAGGCATCCGGGGAGCCGACTTCCAGTGCTTCCAGCAGGCGCGGCCGCG 120
           ||||||||||||||||||||| |||||||||||||||||||||||||| |||||||||
Db   61 GGCGGCATGCGGGGCATCCGGGGCCGACTTCCAGTGCTTCCAGCAGGCGCGGCCGCGGTG 120

Qy   121 GGGCTGGCCGCGCACCTTCGGGGCTTCCTGTGTCGGCGGACCTCTACAGCATC 180
           ||||||| |||||||||||| ||||||||||||||||||||||||||||||||
Db   121 GGGCTGGCGGACCCCAACGGGCCGCAGGCTGCACTGTGAGACGTGGCGGACGGTGTACAGCATC 180

Qy   181 GTGCGCGCGCCGACCGACCCGGGTGCCCGTGTCAACCTCAAGGACGAGGTGCTCTTC 240
           |||||||||||||||||| ||||||||| |||||||||||||| ||||||||| |||
Db   181 GTGCGCGCGTGCCGACCGCGCAGCCGTGCTGCCGTCGTGTCAACCTCAAGGACGAGCTGCTGTTT 240

Qy   241 CCCAGCTGGGAGGGCCTTATTCTCCGGGCTCCGAGGGCCCAGCTGAAGCCCGGGGCCCGCCG 60
           |||||||||||| ||| | |||||||||| ||||||||||||| ||||||||||
Db   241 CCCAGCTGGGAGGGCTCTGTTCTCAGGGTCCGGTTGAAGCCCGGGGGCACGCATC 60

Qy   61 GGCGGCATGCGGAGGCATCCGGGAGCCGACTTCCAGTGCTTCCAGCAGGCGCGGCCGCCGCG 120
           |||||||||| |||||||||| ||||||||||||||||||||||||||||||||| |||||
Db   61 GGCGGCATGCGGGGCATCCGGGCCGGGCCCGACTTCCAGTGCTTCCAGCAGGCGCGGGCCGGGGTG 120

Qy   121 GGGCTGGCCGCGGCACCTTCGGGGCTTCCTGTGTCGGCGGCAGGACCTCTACAGCATC 180
           |||||||| |||||||||||| |||||||||||||||||||||||||||||||||||||
Db   121 GGGCTGGCGGCACCCTTCCTGTCCTGCGTCCTGCCGAGGACCTGTACAGCATC 180

Qy   181 GTGCGCGCGCCCGGTGCCGTGTCAACCTCAAGGACGAGGTGCTCTTC 240
           ||||||||||| ||| ||||||||||||||||||| ||||||| |||
Db   181 GTGCGCGGTGCCGACCGCGCGGCCGCAGCGTCGCCGTGCCATCGTCGTCTGAGGGGACGAGCGCCG 60

Qy   121 GGGCTGGCGGCACCTTCGGGGCTTCCTGTCCTGCGTCCTGCGGCAGGACCTCCAGCAGCCTCTCACGCATC 180
           |||||||||||| || |||| |||||||||||||| ||||||||||| | || | || || | | | | |
Db   121 TTCTCCTTTGACGGCAAGGAGACGTCCTCTGAGGACGTCCTGAGGCACCACCCCGCGGTGTGG 360

RESULT 14

ORIGIN

Query Match   75.3%; Score 418; DB 6; Length 563;
Best Local Similarity 85.3%; Pred. No. 1.9e-46;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy    1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGCCCTGAACAGCCCGCAGCCG  60
        ||||| ||||| |||||||||| |||||||  |||||       ||||||||||| |
Db    6 CACAGCCACCGGCGACTTCCAGCCGGTGCGTCCAGTGCGTTCCAGCCGGTGCCGCCG  65

Qy   61 GGGGGCATCGGCGGGCATCCGGGGAGCGGAGCTTCAGTGCTTCCAGTGCGTTCCAGCCGGTG 120
        ||| |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66 GGGCGCATCGGCGGGCATCCGGGGAGCGGAGCTTCAGTGCTTCCAGTGCGTTCCAGCCGGTG 125

Qy  121 GGGCTGGCGCCGGCACCTTCCGGGCCTTCCTGTCGTCGCGGCTGCAGGACCTTACAGCATC 180
        |||||||||  ||||||||||||||||||||||||||||||||||||||||| ||||||||
Db  126 GGGCTGGCGGGGGCACCTTCCGGGCCTTCCTGTCGTCGCGGCTGCAGGACCTGTACAGCATC 185

Qy  181 GTGCGCGCCGACGGCACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 240
        ||||| ||||  ||||| ||  ||||| |||||||||||| |||||||  |||| ||
Db  186 GTGCGGGCCGTGCGCCGGCAGCCGTGCCCATCGTCAACCTCAAGGACGAGCTGCTGTTT 245

Qy  241 CCCAGCTGGGAGGCCTTATTCTCGGGCTCGTTCTCAGGCTCGAGGGCCCGCATC 300
        |||||||||||||| | |||||||||||||  |||||||||||||||||||||||
Db  246 CCCAGCTGGGAGGCTCGTTCTCAGGCTCGAGGGCTGAGGGCCGAGCCCGGGGGCACGCATC 305

Qy  301 TTCTCTTTGACGGCAGAGATGTCCTGCAGGACTTGGCGGGCCAGGCGGTGTGG 360
        ||||| |||||||||| || |||||||||| ||  |||||||  |||||||||
Db  306 TTCTCCTTTGACGGCAAGGACGTCCTGAGCCGCTGGCGGGCGGCAGGCGGTGTGG 365

Qy  361 CACGGCTCCGACCCCAGCGGGCGCCGCCTGACCGAGAGCGTGGAGGACG 420
        ||||| || ||||||| |||||| ||||||||||||||||| || ||||
Db  366 CATGGCTCGGACCCCAACGCGGCGGCAGCTACCGAGAGCGTGGAGGACG 425

Qy  421 GAGGCCGGCCCCGGCGGCCACCGGCCAGGCCGCGTCGCTGGCGGCCAGGCTGCTCGGAGCAG 480
        |||||||    ||||||||||| |||||||  |||||||||||||||||||||||||||||||
Db  421 GAGGCCCCGGGCCGGCGGCCACCCGGCCAGGCAGGCGTCGCTGGCGGCCAGGCTGCTCGGAGCAG 485

Qy  481 GAGGCCGGCGAGCTGCGCCACGCCTTCGTGGTGTCTGCATCGAGAACAGCGTCATGACC 540
        | |||||||||||||| |||||||||| ||| | |||||| ||||||||| ||||||
Db  481 AGTGCCGCGGAGCTGCCATCACGCCTACATCGTGCTCTGCATTGAGAACAGCTTCATGACT 540

Qy  541 TCCTTC 546
         ||| |
Db  541 GCCTCC 546
```

RESULT 15

ORIGIN

Query Match   75.3%; Score 418; DB 9; Length 786;
Best Local Similarity 85.3%; Pred. No. 1.7e-46;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy    1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGCCCTGAACAGCCCGCAGCCG  60
        ||||| ||||| |||||||||| |||||||  |||||       ||||||||||| |
Db  235 CACACGCACCGGCGACTTCCAGCCGGTGCTCAACCTGGTTGCGCTCAACAGCCCCGTGTCA 294

Qy   61 GGGGGCATGGAGGCATCCGGGGAGCGGAGCTTCAGTGCTTCCAGGCAGGCGGGGCCGCG 120
        |||| |||| |||||| ||||||| |||| |||||||  ||||||||| |||||||||
Db  295 GGGGCATGGGCGGCCGGCATCGCGGCCGGCCGACTTCCAGGTGCTGCGGCAGGCGGGCCGTG 354

Qy  121 GGGCTGGCGGCACCTTCCGGGCCTTCCTGTCGTCGCGGCTGCAGGACCTTACAGCATC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db  355 GGGCTGGCGGCACCTTCCGGGCCTTCCTGTCGTCGCGGCTGCAGGACCTGTACAGCATC 414

Qy  181 GTGCGCGCCGGCACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 240
        ||||| ||||| |  ||  ||||| |||||||||||| ||||||| |||||| |
Db  415 GTGCGGGCCGGTGCGGACCGCAGCGGTGCCATCGTCAACCTCAAGGACGAGCTGCTCTGTTT 474

Qy  241 CCCAGCTGGGAGGCCTTATTCTCGGGCTCGAGCTCTGAGGGCCCGCATC 300
        ||||||||||||||   |  |||||||||| |||||||||||| ||||||
Db  475 CCCAGCTGGGAGGCTGCTCCAGGCTCGAGGGCTGAGGGCCGAGCCCGGGGGCACGCATC 534

Qy  301 TTCTCTTTGACGGCAGAGATGTCCTGCAGACACCCCGGCGCGTGTGG 360
        ||||| |||||||||| || |||||||| ||||||||| | |||||||
Db  535 TTCTCCTTTGACGGCAAGGACGTCCTGAGGCACCCCACCTGGCCCCCAGAAGAGCGGTGTGG 594

Qy  361 CACGGCTTCGACCCCAGGCAGCGCGCCCTGACCGAACAGCTACTCGGAGACG 420
        |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   595 CATGGCTCGGACCCCAACGGGCCGCAGGCTGACCGAGAGCTACTGTGTGAGACGTGGCCGGACG 654

Qy   421 GAGGCCCCGCCGGCGGCCACCGGGGCAGGCGTCGTCGTGCTGCTGGCCGGCGGCTCGTGCTGGAGCAG 480
          |||||||  ||||||  |||||||||||  |||||||| |||||  ||||| ||||| |||
Db   655 GAGGCTCCCTCGGCCACGGGCCAGGCCTCCTCGCTGCTGGTGGGGGGCAGGCTCCTGGGGCAG 714

Qy   481 GAGGCCGCGAGCTGCCGCCACGCCTTCGTGGTGCTCTGCATCGAGAACACAGCGTCATGACC 540
          |||||||||||||  ||||||| ||||||||||| ||||  |||||||||  ||||||||
Db   715 AGTGCCGCGGAGCTGCCATCACGCGCTACATCGTTGCTCTGCATTGAGAACAGCTTCATGACT 774

Qy   541 TCCTTC 546
          |||||
Db   775 GCCTCC 780
```

Search completed: March 29, 2004, 13:38:51
Job time : 2235.23 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: March 29, 2004, 09:08:09 ; Search time 247.424 Seconds
                                      (without alignments)
                                      9529.183 Million cell updates/sec

Title: US-09-938-391-3
Perfect score: 555
Sequence: 1 cacacccacaggacttca........tgacctccttctccaagtag 555

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq_29Jan04:*
              1: geneseqn19806:*
              2: geneseqn1990s:*
              3: geneseqn20008:*
              4: geneseqn2001as:*
              5: geneseqn2001bs:*
              6: geneseqn2002s:*
              7: geneseqn2003as:*
              8: geneseqn2003bs:*
              9: geneseqn2003cs:*
              10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	555	100.0	555	6	AAL46063	Aal46063 Canine en
2	555	100.0	829	6	AAL46062	Aal46062 Canine pr
3	550.4	99.2	552	3	AAZ51309	Aaz51309 Canine an
4	419.6	75.6	1564	6	ABQ76740	Abq76740 DNA encod
5	418	75.3	546	2	AAA27004	Aaa27004 Human end
6	418	75.3	549	3	AAX77719	Aax77719 Human end
7	418	75.3	549	3	AAZ51291	Aaz51291 Human ang
8	418	75.3	549	5	AAA29884	Aaa29884 Human ang
9	418	75.3	549	7	AAS00867	Aas00867 Human ang
10	418	75.3	549	2	ABA00774	Aba00774 Human end
11	418	75.3	552	2	AAX35375	Aax35375 SEQ ID 50
12	418	75.3	552	3	AAA68203	Aaa68203 Human end
13	418	75.3	552	3	AAC62023	Aac62023 Nucleotid
14	418	75.3	552	6	AAC88289	Aac88289 Human end
15	418	75.3	555	4	ABK50685	Abk50685 cDNA enco
16	418	75.3	558	6	AAI66529	Aai66529 Human vas
17	418	75.3	558	6	ABA99261	Aba99261 Human end
18	418	75.3	563	3	AAZ08750	Aaz08750 Human end
19	418	75.3	816	3	AAA64013	Aaa64013 DNA encod
20	418	75.3	900	3	ABK09977	Abk09977 Synthetic
21	418	75.3	968	3	AAC62025	Aac62025 Nucleotid
22	418	75.3	3394	2	AAT84484	Aat84484 Human alp
23	418	75.3	3394	2	AAX78379	Aax78379 Human alp
24	418	75.3	3394	6	ABN95680	Abn95680 Gene #217
25	418	75.3	4551	6	ABN85301	Abn85301 Human col
26	418	75.3	4551	6	ABV94763	Abv94763 Human pan
27	418	75.3	4875	6	ABQ54955	Abq54955 Human ova
28	418	75.3	5408	2	AAX77720	Aax77720 Human col
29	417	75.1	551	6	ABQ81193	Abq81193 Human end
30	415.8	74.9	641	4	AAH79104	Aah79104 Human end
31	415.8	74.9	641	6	AAL55454	Aal55454 Specific
32	415.8	74.9	641	6	ABQ76079	Abq76079 Anticance
33	415.8	74.9	641	6	AAL44000	Aal44000 Reproduct
34	411.6	74.2	555	9	ADE40255	Ade40255 Human NOV
35	410.8	74.0	534	3	AAA27005	Aaa27005 Alternate
36	410.8	74.0	537	5	AAS00868	Aas00868 Human gen
37	409.4	73.8	6462	3	AAA64014	Aaa64014 Nucleotid
38	398.4	71.8	650	6	ABK09978	Abk09978 Synthetic
39	394.2	71.0	552	3	AAZ50398	Aaz50398 Human end
40	375.8	67.7	555	2	AAX79949	Aax79949 Endostati
41	375.8	67.7	555	5	AAX84635	Aax84635 Mouse end
42	375.8	67.7	558	5	AAC88290	Aac88290 Murine en
43	375.8	67.7	558	4	AAD18701	Aad18701 Mouse end
44	375.8	67.7	558	6	ABQ76173	Abq76173 Rat endos
45	375.8	67.7	558	6	ABQ76178	Abq76178 Rat endos

ALIGNMENTS

RESULT 1
AAL46063
ID   AAL46063 standard; cDNA; 555 BP.
XX
AC   AAL46063;
XX
DT   19-JUL-2002 (first entry)
XX
DE   Canine endostatin coding sequence.
XX
KW   Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;
KW   psoriasis; rheumatoid arthritis; retinopathy; macular degeneration;
KW   corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
KW   rubeosis; Osler-Webber Syndrome; myocardial angiogenesis;
KW   plaque neovascularisation; telangiectasia; haemophiliac joints;
KW   angiofibroma; wound granulation; coronary collateral;
KW   cerebral collateral; arteriovenous malformation;
KW   ischaemic limb angiogenesis; diabetic neovascularisation; fracture;
KW   cytostatic; antiarthritic; antipsoriatic; antidiabetic; ophthalmological;
KW   gynaecological; gene; ss.
XX
OS   Canis familiaris.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..555
FT   /tag= a
FT   /product= "endostatin"
FT   /partial
FT   /note= "no start codon"
XX
PN   EP191036-A2.
XX
PD   27-MAR-2002.
XX
PF   24-AUG-2001; 2001EP-00307224.
XX
PR   25-AUG-2000; 2000US-0227924P.
XX
PA   (PFIZ) PFIZER PROD INC.
XX
PI   Sheppard MG, Tong X;
XX
DR   WPI; 2002-354068/39.
DR   P-PSDB; AAO17430.
XX
PT   An isolated nucleic acid molecule for the treatment of angiogenesis-

PT Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;
PT psoriasis; rheumatoid arthritis; retinopathy; macular degeneration;
XX corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
KW rubeosis; Osler-Webber Syndrome; myocardial angiogenesis;
KW plaque neovascularisation; telangiectasia; haemophiliac joints;
KW angiofibroma; wound granulation; coronary collateral;
KW cerebral collateral; arteriovenous malformation;
KW ischaemic limb angiogenesis; diabetic neovascularisation; fracture;
KW cytostatic; antiarthritic; antipsoriatic; antidiabetic; ophthalmological;
KW gynaecological; gene; ss.
OS Canis familiaris.
XX
FH Key   Location/Qualifiers
FT CDS   1..693
FT   /*tag= a
FT   /product= "pro-endostatin"
FT   /partial
FT   /note= "no start codon"
XX
PN EP1191036-A2.
XX
PD 27-MAR-2002.
XX
PF 24-AUG-2001; 2001EP-00307224.
XX
PR 25-AUG-2000; 2000US-0227924P.
XX
PA (PFIZ) PFIZER PROD INC.
XX
PI Sheppard MG, Tong X;
XX
DR WPI; 2002-354068/39.
DR P-PSDB; AAO17429.
XX
PT An isolated nucleic acid molecule for the treatment of angiogenesis-
PT related disorder, such as cancers or diabetic retinopathy, encodes an
PT endostatin protein.
XX
PS Claim 2; Fig 2; 56pp; English.
XX
CC The present invention provides the protein and coding sequences of canine
CC pro-endostatin and endostatin. The sequences can be used in the treatment
CC and diagnosis of angiogenesis related disorders, including cancer,
CC rheumatoid arthritis, psoriasis, retinopathy, macular degeneration,
CC corneal graft rejection, neovascular glaucoma, retrolental fibroplasia,
CC rubeosis, Osler-Webber Syndrome, myocardial angiogenesis, plaque
CC neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,
CC wound granulation, coronary collaterals, cerebral collaterals,
CC arteriovenous malformations, ischaemic limb angiogenesis, diabetic
CC neovascularisation, and fractures. The present sequence is the canine pro
CC -endostatin coding sequence
XX
SQ Sequence 829 BP; 124 A; 314 C; 278 G; 113 T; 0 U; 0 Other;

Query Match   100.0%; Score 555; DB 6; Length 829;
Best Local Similarity 100.0%; Pred. No. 1.7e-83;
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   139 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 198

Qy   61 GGCGGCATGCGAGGCATCCGGGAGCCGACTTCCAGTGCTTCCAGCAGGCGCGCCGGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   199 GGCGGCATGCGAGGCATCCGGGAGCCGACTTCCAGTGCTTCCAGCAGGCGCGCCGGG 258

Qy   121 GGGCTGGCCGGCACCTTCCGGGCCTTCCGTGTCGCGGCTGCAGGACCTTCTACAGCATC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   259 GGGCTGGCCGGCACCTTCCGGGCCTTCCGTGTCGCGGCTGCAGGACCTTCTACAGCATC 318

Qy   181 GTGCCCGGCGCCGACCGGAC 240

```
PT                      as rheumatoid arthritis, tumors and macular degeneration.
XX
PS   Example 8; Page 58-59; 68pp; English.
XX
CC   The patent discloses a DNA molecule encoding a fusion protein comprising
CC   a signal sequence, an immunoglobulin Fc region, and an angiogenesis
CC   inhibitor selected from angiostatin, endostatin, a plasminogen fragment
CC   having angiostatin activity, a collagen XVIII fragment having endostatin
CC   activity, or combinations of them. The fusion protein (immunofusin) is
CC   used to inhibit angiogenesis and to treat diseases or conditions mediated
CC   by angiogenesis. Conditions that may be treated include solid tumours,
CC   blood born tumours, tumour metastasis, benign tumours including
CC   haemangiomas, acoustic neuroma, neurofibromas, trachomas and pyrogenic
CC   granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
CC   e.g. diabetic retinopathy, retinopathy of prematurity, macular
CC   degeneration, corneal graft rejection, neovascular glaucoma, retrolental
CC   fibroplasia, rubeosis and Osler-Webber syndrome; myocardial angiogenesis,
CC   plaque neovascularisation, telangiectasia, haemophiliac joints'
CC   angiofibroma, wound granulation, and excessive or abnormal stimulation of
CC   endothelial cells, intestinal cells, atherosclerosis, sclerodermal and
CC   hypertrophic scars, i.e. keloid scars. The DNA constructs may be used in
CC   gene therapy. The present sequence is a DNA encoding canine endostatin
CC   used in the construction of immunofusin containing canine immunoglobulin
CC   Fc fragment
XX
SQ   Sequence 552 BP; 77 A; 204 C; 190 G; 81 T; 0 U; 0 Other;

Query Match             99.2%;  Score 550.4;  DB 3;  Length 552;
Best Local Similarity   99.8%;  Pred. No. 1e-82;
Matches  551;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   0;

Qy         1 CACACCACCAGGACTTCCAGTCGGTGGCTGCACCTGGTGCCCTGAACAGCCCAGCCG  60
             | |||| |||||||||||||| ||||| || ||||||||||||||||||||||||||
Db         1 CACACCCACCAGGACTTCCAGCCGGTGCTGCACCTGGTGCCCTGAACAGCCCAGCCG  60

Qy        61 GGGCTGCATGCGAGGCATCCGGGAGGCCGGACTTCCAGTGCTTCCAGTCGCCGCCGCCG 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 GGGCTGCATGCGAGGCATCCGGGAGGCCGGACTTCCAGTGCTTCCAGTCGCCGCCGCCG 120

Qy       121 GGGCTGCCGGCCACCTTCCGGGCCTTCCTGCGTCGTGCGGCTGCAGGACACCTTCTACAGCATC 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 GGGCTGCCGGCCACCTTCCGGGCCTTCCTGCGTCGTGCGGCTGCAGGACCTTCTACAGCATC 180

Qy       181 GTGCCGCCGGCGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GTGCCGCCGGCGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 240

Qy       241 CCCAGCTGGGAGGCCTTATTCTCGGGGCTCCGAGGGGCCAGCTGAAGCCCGGGGCCCGCATC 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 CCCAGCTGGGAGGCCTTATTCTCGGGGCTCCGAGGGGCCAGCTGAAGCCCGGGGCCCGCATC 300

Qy       301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGACACCCCGGGAGAGCGGTGTGG 360
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGACACCCCGGGAGAGCGGTGTGG 360

Qy       361 CACGGCTCCGACCCCAGCGGCGGCCGCCGCTGACCGAGCAGTACTGCGAGACGTGGCGGACG 420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 CACGGCTCCGACCCCAGCGGCGGCCGCCGCTGACCGAGCAGTACTGCGAGACGTGGCGGACG 420

Qy       421 GAGGCCCCGGCCACCGGGCCAGGGCCTGCGTCCTCGGCGGGCAGGCTGCTGAGCAG 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 GAGGCCCCGGCCACCGGGCCAGGGCCTGCGTCGTCGGCGGGCAGGCTGCTGGAGCAG 480

Qy       481 GAGGCCGCCGAGTGCGCGCCCAGCCCTTCGTGGTGGCTCCGATCGGAGAACAGCGTCATGACC 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 GAGGCCGCCGAGTGCGCGCCCAGCCCTTCGTGGTGGCTCCGATCGGAGAACAGCGTCATGACC 540

Qy       541 TCCTTCTCCAAG 552
             ||||||||||||
Db       541 TCCTTCTCCAAG 552

RESULT 4
```

```
Db       319 GTGGCGCGGCCGACCCCAGCGGCACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 378
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       241 CCCAGCTGGGAGGCCTTATTCTCGGGGCTCCGAGGGGCCAGCTGAAGCCCGGGGCCCGCATC 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       379 CCCAGCTGGGAGGCCTTATTCTCGGGGCTCCGAGGGGCCAGCTGAAGCCCGGGGCCCGCATC 438
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGACACCCCGGGAGAGCGGTGTGG 360
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       439 TTCTCTTTCGACGGCAGAGATGTCCTGCAGACACCCCGGGAGAGCGGTGTGG 498
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       361 CACGGCTCCGACCCCAGCGGCGGCCGCCGCTGACCGAGCAGTACTGCGAGACGTGGCGGACG 420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       499 CACGGCTCCGACCCCAGCGGCGGCCGCCGCTGACCGAGCAGTACTGCGAGACGTGGCGGACG 558
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       421 GAGGCCCCGGCCACCGGGCCAGGGCCTGCGTCGTCGGCGGGCAGGCTGCTGGAGCAG 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       559 GAGGCCCCGGCCACCGGGCCAGGGCCTGCGTCGTCGGCGGGCAGGCTGCTGGAGCAG 618
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       481 GAGGCCGCCGAGTGCGCGCCCAGCCCTTCGTGGTGGCTCCGATCGGAGAACAGCGTCATGACC 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       619 GAGGCCGCCGAGTGCGCGCCCAGCCCTTCGTGGTGGCTCCGATCGGAGAACAGCGTCATGACC 678
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       541 TCCTTCTCCAAGTAG 555
             |||||||||||||||
Db       679 TCCTTCTCCAAGTAG 693

RESULT 3
AA251309
ID   AA251309 standard; DNA; 552 BP.
XX
AC   AA251309;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Canine angiogenesis inhibitor, endostatin DNA.
XX
KW   Canine; immunoglobulin Fc fragment; endostatin; immunofusin;
KW   angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;
KW   antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;
KW   vasotropic; vulnerary; treatment; antiarteriosclerosis; tumour;
KW   metastais; atherosclerosis; psoriasis; rheumatoid arthritis;
KW   ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW   myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW   wound granulation; keloid scar; gene therapy; ds.
XX
OS   Canis familiaris.
XX
FH   Key            Location/Qualifiers
FT   CDS            1..552
FT                  /*tag= a
FT                  /product= "Endostatin"
FT                  /note= "Does not include stop codon"
FT                  /partial
XX
PN   WO200011033-A2.
XX
PD   02-MAR-2000.
XX
PF   25-AUG-1999; 99WO-US019329.
XX
PR   25-AUG-1998; 98US-0097883P.
XX
PA   (LEXI-) LEXINGEN PHARM CORP.
XX
PI   Lo K, Li Y, Gillies SD;
XX
DR   WPI; 2000-237616/20.
DR   P-PSDB; AAY70265.
XX
PT   Novel fusion protein of angiostatin or endostatin and an immunoglobulin
PT   FC region, useful for treating conditions mediated by angiogenesis, such
```

```
ABQ76740
ID   ABQ76740 standard; DNA; 1564 BP.
XX
AC   ABQ76740;
XX
DT   03-MAR-2003  (first entry)
XX
DE   DNA encoding human Endostatin/IgG1Fc fusion construct.
XX
KW   Human; endostatin; IgG1Fc; tumour; vascular endothelial proliferation;
KW   vascular endothelial cytopoiesis inhibiting factor; inhibitor;
KW   fusion construct; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key           Location/Qualifiers
FT   CDS           1..1527
FT                 /tag= a
FT                 /product= "endostatin/IgG1Fc construct"
FT                 /transl_except= (pos58..60,aa:Ala)
FT                 /transl_except= (pos:811..816,aa:GGGGSGG)
XX
PN   CN1354186-A.
XX
PD   19-JUN-2002.
XX
PF   30-NOV-2000; 2000CN-00123347.
XX
PR   30-NOV-2000; 2000CN-00123347.
XX
PA   (LIAO-) LIAONING WEIXING BIOLOGICAL PROD INST CO.
XX
PI   Chen L, Li Z, Liu Q;
XX
DR   WPI; 2002-751441/82.
DR   P-PSDB; ABG73586.
XX
PT   Preparation of recombinant human vascular endothelial cytopoiesis
PT   suppressor factor with human IgG1Fc fragment molecular structure and
PT   application of its product.
XX
PS   Disclosure; Page 7-8 (Disclosure); 12pp; Chinese.
XX
CC   This invention describes a novel method for the preparation of
CC   recombinant human vascular endothelial cytopoiesis inhibiting factor with
CC   human IgG1Fc fragment molecular structure and its product application.
CC   The novel factor is derived from endostatin (using PCR to screen a human
CC   foetal kidney cell cDNA library) and human IgG1Fc. The product of the
CC   invention can specifically inhibit tumour vascular endothelial
CC   proliferation and can be used for curing several tumour types. This
CC   sequence encodes a fusion construct composed of human endostatin and
CC   human IgG1Fc, described in the disclosure of the invention
XX
SQ   Sequence 1564 BP; 346 A; 491 C; 436 G; 291 T; 0 U; 0 Other;

Query Match   75.6%;   Score 419.6;  DB 6;  Length 1564;
Best Local Similarity 84.8%;  Pred. No. 4.3e-61;
Matches   470;  Conservative   0;  Mismatches   84;  Indels   0;  Gaps   0;

Qy       1 CACACCCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCCCAGCCG  60
           |||  ||||||||||||||||||||||||||||||||| ||||| |  ||||||||||||
Db     262 CACAGCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTTGCGCTCAACACGCCCCTGTCA 321

Qy      61 GGCGGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGGCGCGCGGCCG 120
           ||||||||||||||||||||||||||||||||||| |||| |  ||||||||||||||||
Db     322 GGCGGGCATGCGGGGCATCCGGGGCGCCGGGGACTTCCAGCGCTGCTTCCAGCAGGCCGTG 381

Qy     121 GGGCTGCCGCCGACCTTCCGGGACCTTCTCGCCGTCGTCGTCGGGCTGCCAGGACCATC  180
           ||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db     382 GGGCTGCCGGGCACCTTCCGGCCGCCCTCCTGTCCTGCGCCGGCGACCTGTACAGCATC 441

Qy     181 GTGCGCCCGCGCCCGCACCGGACCGCACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 240
```

```
Db   442 GTGCGCCGTGCCGACGCGCCGCAGCCGTGCCCATCGTCAACCTCAAGGACGACGAGCTGCTGTGTT 501
             |||||||  |||||| |||||||||||||  |||||  ||||||||||||||||||||||||||

Qy   241 CCCAGCTGGGAGGCGTCCGAGGCGGCAGCTGAGCGGGGCGGCCCCGATC 300
             |||||||||||||  |||||||||| ||||||||||||||||||||||
Db   502 CCCAGCTGGGAGGCGTCTCGTTCTCAGGCTCGAGGGTCCGTGAGGCCCCGGAGCCCGGGCACGCATC 561

Qy   301 TTCTCTTCGACGGCAGAGATGTCCTGCGAGCACCCGCATCGGCCCCCGGAAGAGCGTGTGG 360
             |||||||||| || |  || |||| ||||||||||||||||  ||||||||| ||||||||
Db   562 TTCTCCTTTGACGGCAAGAGACGTCCTGAGGCACCCACCTGGCCCCACTACTTCCGACACGGTGTGG 621

Qy   361 CACGGCTCCGACCCCAGCGGCGCGGGCGGCCGCCAGGCGCTGACCAGCTACTCGAGACTTGGCCGACG 420
             ||||||||| |||||||||||||||  ||||  ||| |||||||||||||||| ||||||||||||
Db   622 CATGGCTCGGACCCCAACGGGCGACCCCAGCGGCGCGGGCGGCCGCCAGGCGCTGACCAGCTACTGTGACGCTACTGTGACGGCGCGGACG 681

Qy   421 GAGGCCCGGCGGCGCACCGGCAGGCGTCGTCCGTCGGCCCTGTCTGCCGGCGTGCGGAGCAG 480
             ||||||||||||||| |||| |||||||| |||||  |||| |||||||||||| ||||||
Db   682 GAGGCTCCCTCGGCCACGGCGCCGCCTCCTGGGGGCAGGCTCCTGGGGCAG 741

Qy   481 GAGGGCGGCGAGCTGCGCCACCGCCCTTCGTGGTGCTTCGATCGAGAACAGCGTCATGACC 540
             |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   742 AGTGGCGCGGAGCTGCCATCACGCGCTGCTCTGCTGTCTGCATTGAGAACAGCTTCATGACT 801

Qy   541 TCCTTCTCCAAGTA 554
             |||  ||||||||
Db   802 GCCTCCAAGTAGTA 815
```

```
RESULT 5
AAA27004
ID   AAA27004 standard; DNA; 546 BP.
XX
AC   AAA27004;
XX
DT   11-AUG-2000  (first entry)
XX
DE   Human endostatin gene.
XX
KW   Human; endothelial cell proliferation inhibitor; collagen XVIII;
KW   angiogenesis inhibitor; anti-tumour; cytostatic; antipsoriatic;
KW   vasotropic; dermatological; ophthalmological; vulnerary;
KW   antiarteriosclerotic; antidiabetic; haemostatic; contraceptive;
KW   ocular angiogenic disease; atherosclerosis; scleroderma;
KW   myocardial angiogenesis; telangiectasia; angiofibroma;
KW   wound granulation, ds.
XX
OS   Homo sapiens.
XX
FH   Key           Location/Qualifiers
FT   CDS           1..546
FT                 /tag= a
FT                 /product= "Endostatin protein"
XX
PN   WO200026368-A2.
XX
PD   11-MAY-2000.
XX
PF   01-NOV-1999;  99WO-US025605.
XX
PR   30-OCT-1998;  98US-0106343P.
PR   20-MAY-1999;  99US-00315689.
XX
PA   (CHIL-) CHILDRENS MEDICAL CENT.
XX
PI   O'reilly MS,  Folkman MJ;
XX
DR   WPI; 2000-365617/31.
DR   P-PSDB; AAY94323.
XX
PT   Novel endostatin capable of inhibiting endothelial cell proliferation and
PT   angiogenesis, useful for treating angiogenesis-dependent cancers and as
PT   birth control agents.
XX
```

PS Claim 10; Page 39; 68pp; English.
XX
CC The present sequence encodes an endostatin protein which is the carboxy
CC terminal protein of human collagen XVIII. Recombinant mouse endostatin
CC (20 mg/kg) was administered subcutaneously to mice implanted with Lewis
CC lung carcinomas. There was tumour mass regression non-detectable levels
CC after 12 days of therapy due to the angiogenesis inhibitory activity of
CC endostatin. Thus the protein is useful for treatment of angiogenesis-
CC dependent cancers. The polynucleotide and polypeptide sequences of this
CC endostatin are useful for treating and diagnosis of tumours, ocular
CC angiogenic diseases, Osler-Webber syndrome, myocardial angiogenesis,
CC plaque neovascularisation, telangiectasia, haemophiliac joints,
CC angiofibroma and wound granulation, for treatment of diseases related to
CC excessive or abnormal stimulation of endothelial cells e.g. intestinal
CC adhesions, atherosclerosis, scleroderma. The protein may also be useful
CC as a birth control agent by reducing or preventing uterine
CC vascularisation. The gene for endostatin may be isolated from cells or
CC tissue that express high levels of endostatin, eg. tumour cells, by
CC generating cDNA from mRNA using reverse transcriptase and then amplifying
CC the DNA sequence
XX
SQ Sequence 546 BP; 80 A; 196 C; 177 G; 93 T; 0 U; 0 Other;

Query Match   75.3%; Score 418; DB 3; Length 546;
Best Local Similarity 85.3%; Pred. No. 8.9e-61;
Matches 466; Conservative   0; Mismatches 80; Indels   0; Gaps   0;

Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG   60
        |||||| |||| ||||||||||| | ||| |||||||| ||| ||||
Db   1 CACAGCCACCGGACTTCCAGCCGCGGCATCCGGGGGAGGGCCAGCCGGGCGGGGCCGTG   60

Qy   61 GGCGGCATGCGGAGGCATCCGGGGGAGGGACTTCCAGTGCTTCCAGCTGCTGCGCGCCG 120
        |||||||||||||| ||||||| ||| || ||||||||||||||||||||||
Db   61 GGCGGCATGCGGGGGCATCCGGGGGCCGGACTTCCAGTGCTTCCAGCTGCTCCAGCGGCG 120

Qy 121 GGGCTGCGCCGGCACCTTCCGGGCCTTCCGTGTCGCGGCTGAGGACCTCTACAGCATC 180
        ||||||||||||| ||||| |||||||||||||| |||||||||||||||||
Db 121 GGGCTGGGGGGCACCTTCCGGGCCTTCCGTGTCCTGCGGCTGAGGACCTCTACAGCATC 180

Qy 181 GTGCGGCCGCGCCACCGGGGTGCCCGTGGACGAGGTGCTCTTC 240
        |||||| |||||| ||| ||||||||| ||| ||||||
Db 181 GTGCGGCCGTGCCGACCGCGAGCCGGTGCCCATCGCGGGGCCGCCGCAGGCCGCACGGCATC 240

Qy 241 CCCAGCTGGGAGGCCTTATTCTCGGGGCTCCGAGGGCCAGCTGAAGCCCGGGCCCGCATC 300
        |||||||||||||| ||||||| ||||||| |||| |||||||||||
Db 241 CCCAGCTGGGAGGCTCTGTTCTCCAGGCTCGAGAGCTCGAGGGCCACCCGGGCACGGCATC 300

Qy 301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCGGCCTGGCCCGGCGTGTGG 360
        |||||| |||||| ||| || ||| || ||||| |||||||||||||||
Db 301 TTCTCCTTTGACGGCAAGGACGTCCTCGAGGCACCCGGCCTGGCCCGGCGTGTGG 360

Qy 361 CACGGCTCCGACACCCCAGCGGCGCCGCGCCTGACCGACAGCTACTGCGGACACG 420
        |||||||||||| |||| |||||||||||| ||||| || ||||||||||||
Db 361 CATGGCTCGGACACCCCAACGGGCGCAGGCTGGGCCCGGAGACGGTGCGGCGACG 420

Qy 421 GAGGCCCCGGCGGCCACCGGGCAGGCGTCGTCGCCTGCTGGCGGCAGGCTGCTGAGCAG 480
        ||||||||||| |||||||||||||||||||||||||| |||||||||||||
Db 421 GAGGGCTCCTCGGCCACAGGGCCAGGCGTCCTGGGGCGGCAGGCTCCTGGGGCAG 480

Qy 481 GAGGCGGCGAGCTGGCCACGCCGCCACGCCGCCTTCGTGGTGCTCGCATGCAGCAGGTCATGACC 540
        |||||||| |||||||| |||| |||||||||||||
Db 481 AGTGCGCGGAGCTGCCATCACGCCTACATCGTGCTCGCATTGCGCATTGAACACGCTTCATGACT 540

Qy 541 TCCTTC 546
        | |||
Db 541 GCCTCC 546

RESULT 6
AAX77719
ID AAX77719 standard; DNA; 549 BP.
XX
AC AAX77719;
XX
DT 10-AUG-1999 (first entry)
XX
DE Human endostatin DNA coding region fragment.
XX
KW Plasminogen; human; angiostatin; endostatin; gene therapy; vector;
KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW tumour growth; solid tumour; diabetic retinopathy; retina; ss.
XX
OS Homo sapiens.
XX
PN WO9926480-A1.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-US024950.
XX
PR 20-NOV-1997; 97US-00975424.
XX
PA (GENE-) GENETIX PHARM INC.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Lebouich P, Pawliuk RJ, Bachelot T;
XX
DR WPI; 1999-357696/30.
DR P-PSDB; AAY08693.
XX
PT Anti-angiogenic gene therapy vectors.
XX
PS Disclosure; Page 74; 83pp; English.
XX
CC This invention describes a novel viral gene therapy vector comprising a
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen from
CC human or murine angiostatin, human or murine endostatin and angiogenesis-
CC inhibiting fusions and fragments, where the viral vector is sufficiently
CC attenuated for use in human gene therapy. The products of the invention
CC have anti-angiogenic, cytostatic, anti-diabetic and ophthalmological
CC activity. The vector is used in gene therapy for inhibiting tumour growth
CC in humans harbouring a solid tumour. The vector expresses an anti-
CC angiogenic polypeptide. An additional use comprises treatment of diabetic
CC retinopathy, where the anti-angiogenic polypeptide inhibits angiogenesis
CC in the vicinity of the retina. The vector is administered to cells ex
CC vivo and then administered to the patient
XX
SQ Sequence 549 BP; 82 A; 196 C; 178 G; 93 T; 0 U; 0 Other;

Query Match   75.3%; Score 418; DB 2; Length 549;
Best Local Similarity 85.3%; Pred. No. 8.9e-61;
Matches 466; Conservative   0; Mismatches 80; Indels   0; Gaps   0;

Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG   60
        |||| |||||| |||||||||| ||| ||| |||||| |||| ||||||
Db   1 CACAGCCACCGGACTTCCAGCTGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA   60

Qy   61 GGCGGCATGCGGAGGCATCCGGGGGAGGGACTTCCAGCGGCGGCGCGCGG 120
        ||||||||||||| ||||||||||||| |||||||||||||||||||
Db   61 GGCGGCATGCGGGGGCATCCGGGGGCCGGCCAGCTTCCAGCGGCGGCGCCGGTG 120

Qy 121 GGGCTGCGCCGGCACCTTCCTGTGTCGCGGCTGAGGACCTCAGGACCATC 180
        |||||||||||| |||||||| ||||||||| ||||||||||||||
Db 121 GGGCTGGCGGGGCACCTTCCGTGTCGTCCGCGCCTTCCAGCAGGACCTACAGCATC 180

Qy 181 GTGCGGCCGCGCCACCGGGGTGCCCGTGGACCTGTGCGCGGCGCCGTGTCAGCGATC 240
        ||||||||||| ||| ||||||||| ||| |||| ||||||||||||
Db 181 GTGCGGCCGGGCCACCTGGGCTCAACCTCAGGGTGCCCGTGGCCCATCGCGGGGCCGCCGCAGCCTCTCCAGGACCATC 240

Qy 241 GTGGCGCGGCGCGCAGCCACGCGGGGTGCCCGTGGGCCCTGAAGAGCCACGCGGTGTGG 300
        |||||||||||| ||| || ||||| |||||| |||||||
Db 241 GTGGCGCGGCGCGCGAGCCGGTGCCCATCGCGGGGCCGCCGCAGGCCGCACGGCATC 300

Qy 301 TTCTCTTTCGACGGCAGGAGATGTCCGACACCCCTATTCTCGGGCTCCGAGGGCCCGCATC 360
        |||||| ||||||| |||||| ||||| |||||| |||||| |||||
Db 301 CCCAGCTGGGAGCTCTGTTCTCCAGGCTCGAGGGCTCGGGGCCCGGCCCGGGGCACGCATC 300

```
CC        by angiogenesis. Conditions that may be treated include solid tumours,
CC        blood born tumours, tumour metastasis, benign tumours including
CC        haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic
CC        granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
CC        e.g. diabetic retinopathy, retinopathy of prematurity, macular
CC        degeneration, corneal graft rejection, neovascular glaucoma, retrolental
CC        fibroplasia, rubeosis and Osler-Webber syndrome; myocardial angiogenesis,
CC        plaque neovascularisation, telangiectasia, haemophiliac joints'
CC        angiofibroma, wound granulation, and excessive or abnormal stimulation of
CC        endothelial cells, intestinal cells, atherosclerosis, sclerodermal and
CC        hypertrophic scars, i.e. keloid scars. The DNA constructs may be used in
CC        gene therapy. The present sequence is a cDNA encoding human endostatin
CC        used in the construction of immunofusin containing human immunoglobulin
CC        gamma (IgG) Fc fragment. Note: This sequence is stated in claim 12 as
CC        being amino acid sequence of plasminogen fragment, however, the rest of
CC        the specification refers to this sequence as being nucleotide sequence of
CC        human endostatin
XX
SQ    Sequence 549 BP; 82 A; 196 C; 178 G; 93 T; 0 U; 0 Other;

Query Match        75.3%;  Score 418;  DB 3;  Length 549;
Best Local Similarity  85.3%;  Pred. No. 8.9e-61;
Matches  466;  Conservative  0;  Mismatches  80;  Indels  0;  Gaps  0;

Qy      1  CACACCCACCAGGACTTCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG   60
           |||||||||| ||||||||||||||||| ||||||||||||||||||| ||||||||||
Db      1  CACAGCCACCCGGACTTCCAGCCGGTGCTCCACCTGGTTGCGCTCAACAGCCCCTGTCA   60

Qy     61  GGCGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCAGCAGCGGCGCGCGCG   120
           ||||||||| |||||||||||| ||| |||||||||||||||||||||||||| |||||
Db     61  GGCGGCATGCGGGGCATCCGGGGCCGGACTTCCAGTGCTTCCAGCAGCGGCGCGGCCGTG   120

Qy    121  GGGCTGGCCGGCACCTTCCTGTCGTCGCGGCTGCAGGACCTTCTACAGCATC   180
           |||| |||||||||||||||||||||| ||||||||||||||||| |||||||
Db    121  GGGCTGCGGGGCACCTTCCTGCGTCCTCGCGCCTGCAGGACCTGTACAGCATC   180

Qy    181  GTGCGCCGCGCGACCGCACCGGGGTGCCCGTGCCATCGTCAACCTCAAGGACGAGGTGCTCTTC   240
           ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    181  GTGCGGCGTGCCGACCGCACCGGGCTCCAGAGCGTGCCATCGTCAACCTCAAGGACGAGCTGCTGTT   240

Qy    241  CCCAGCTGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGCCCCCATC   300
           |||||||||||||||||||||| ||||||||||||||||||||||||||| |||||||
Db    241  CCCAGCTGGGAGGCCTCTGTTCTCAGGGTGAGGGTCCGGTGAAGCCCGGGCACGCATC   300

Qy    301  TTCTCTTCGACGCGGCAGAGATGTCCTGCAGCACCCGCCTGGCCCCGGAGAAGAGCGTCTGG   360
           |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db    301  TTCTCCTTTGACGCGGCAAGAGACGTCCTGAGCGCACCCACCTGGCCCCACCTGGCCCCAGAAGAGCGTGTGG   360

Qy    361  CACGGCTCCGACCCCAGCCGGCGGCGCCTGACCACGCTACTGCCGGACG   420
           ||||||| |||||||||| |||| ||||||||||||||||||||||||||
Db    361  CATGGCTCGGACCCCAACGGGCGGCCGCGAGAGCTACTGTGAGACGTGGCCGACG   420

Qy    421  GAGGCCCGGCGGCCACCGGCAGGCGTCGTCGCTGGTGGCGGCAGCGTGCTGGAGCAG   480
           |||||| ||||||||||||||||||||||||||||| |||||||||||| |||||||
Db    421  GAGGCTCCCTGCGCCACGGGCCAGGCCTCCTGCTGCGCTGCGTGGGGGGCAGGCTCCTGGGGCAG   480

Qy    481  GAGGCCGGAGCTGCGCCACGCCTTCGTGGTGCTCTCGATCGAGAACAGCGTCATGACC   540
           ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    481  AGTGCGCGGAGCTGCCATCACCGCCTACATCGTGCTTGCATTGAGAACCAGCTTCATGACT   540

Qy    541  TCCTTC  546
           |||||||
Db    541  GCCTCC  546

RESULT 7
AAZ51291
ID   AAZ51291 standard; cDNA; 549 BP.
XX
AC   AAZ51291;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Human angiogenesis inhibitor, endostatin cDNA.
XX
KW   Human; immunoglobulin gamma Fc fragment; endostatin; immunofusin;
KW   angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;
KW   antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;
KW   vasotropic; vulnerary; treatment; antiarteriosclerosis; tumour;
KW   metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
KW   ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW   myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW   wound granulation; keloid scar; gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key              Location/Qualifiers
FT   CDS              1..549
FT                    /*tag= a
FT                    /product= "Endostatin"
FT                    /note= "Does not include stop codon"
FT                    /partial
XX
PN   WO20001I033-A2.
XX
PD   02-MAR-2000.
XX
PF   25-AUG-1999;  99WO-US019329.
XX
PR   25-AUG-1998;  98US-0097883P.
XX
PA   (LEXI-) LEXINGEN PHARM CORP.
XX
PI   Lo K, Li Y, Gillies SD;
XX
DR   WPI; 2000-237616/20.
DR   P-PSDB; AAY70252.
XX
PT   Novel fusion protein of angiostatin or endostatin and an immunoglobulin
PT   FC region, useful for treating conditions mediated by angiogenesis, such
PT   as rheumatoid arthritis, tumors and macular degeneration.
XX
PS   Claim 12; Page 40-41; 68pp; English.
XX
CC   The patent discloses a DNA molecule encoding a fusion protein comprising
CC   a signal sequence, an immunoglobulin Fc region, and an angiogenesis
CC   inhibitor selected from angiostatin, endostatin, a plasminogen fragment
CC   having angiostatin activity, a collagen XVIII fragment having endostatin
CC   activity, or combinations of them. The fusion protein (immunofusin) is
CC   used to inhibit angiogenesis and to treat diseases or conditions mediated
```

RESULT 8
AAA29984
ID   AAA29984 standard; cDNA; 549 BP.
XX
AC   AAA29984;
XX
DT   22-AUG-2000 (first entry)
XX

DE   Human angiogenesis inhibiting factor 1 encoding cDNA.
XX
KW   Human; angiogenesis inhibiting factor 1; IAF-1; tumour; antibody;
KW   abnormal vessel disease; ss.
XX
OS   Homo sapiens.
XX
PN   CN1244536-A.
XX
PD   16-FEB-2000.
XX
PF   10-AUG-1998; 98CN-00117150.
XX
PR   10-AUG-1998; 98CN-00117150.
XX
PA   (ONCO-) INST ONCOLOGY UNDER TUMOR HOSPITAL CHINE.
XX
PI   Yang Z, Guo W;
XX
DR   WPI; 2000-388168/34.
DR   P-PSDB; AAY90771.
XX
PT   Angiogenesis inhibiting factor 1 and its derivative useful for treating
PT   tumors.
XX
PS   Example 1; Fig 5; 41pp; Chinese.
XX
CC   The present sequence encodes an angiogenesis inhibiting factor (I),
CC   designated IAF-1. The present invention also describes: (1) preparation
CC   of (I) and its derivative; (2) an IAF binding acceptor and its
CC   preparation; and (3) an IAF antibody. (I) is useful for preparing new
CC   biological preparations for effectively treating various tumours and
CC   abnormal-vessel diseases. The IAF antibody is preferably a polyclonal
CC   antibody, mosaic antibody, single stranded antibody and human originated
CC   antibody
XX
SQ   Sequence 549 BP; 82 A; 196 C; 178 G; 93 T; 0 U; 0 Other;

Query Match   75.3%; Score 418; DB 3; Length 549;
Best Local Similarity   85.3%; Pred. No. 8.9e-61;
Matches 466; Conservative   0; Mismatches   80; Indels   0; Gaps   0;

Qy   1 CACACCCACCAGGACTTCCAGTCGTGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 60
        ||||| || || ||||||||||| || ||| |||||||| | ||||||||||| |
Db   1 CACAGCCACCGCGACTTCCAGCCGTGCTCCAGTGGTGCCGCTCAACAGCCCCCCTGTCA 60

Qy   61 GGGGCATGCGAGGCATCGGGGAGGCGGACTTCCAGTGCTTCCAGCAGCGGCGCGGCCGCG 120
        |||||||||| |||||| || ||| |||||||||||||||||||||||||| ||| ||
Db   61 GGGGCATGCGGGGGCATCCCGCCGGGCCGACTGCTTCCAGTGCTTCCAGCAGCGGCGGCCGTG 120

Qy 121 GGGCTGGCCGGCACCTTCCGGGCCTTCCTGCGTCGCGGCGTCCAGGACCTTCACAGCATC 180
        ||||||||||||||||||||||||||||||| || ||||||||||||||| |||||||||
Db 121 GGGCTGGCCGGCACCTTCCGGGCCTTCCTGTCCTCGCGGCGTCCAGGACCTTGTACAGCATC 180

Qy 181 GTGCGCGCGCGAGCCGCCACCGGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 240
        |||||||||||||| ||||||||||| |||||||| ||||||||||||||||| ||||
Db 181 GTGCGCGCGTGCCGAGCCGACCGGGCCGCAGCGGCGTCAACCTCAAGGACGAGCTGCTGTTT 240

Qy 241 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGGGCCGCTCCGAGGGCCCAGCTGAAGCCCGCATC 300
        |||||||||||||| | | |||||||||||||| ||||||| |||||| ||||||||||||
Db 241 CCCAGCTGGGAGGGCTCTGTTCTCCAGGCTCTGGAGGGTCCGCTGAGGCCTGAAGCCCGGGCCACGCATC 300

Qy 301 TTCTCTTCGACGGCAGCAGAGATGTCCTGCAGCACCCGCGTGTGG 360
        ||||| |||||||||||||| | ||||||||||||||||||||
Db 301 TTCTCCTTTGACGGCAAGGACGTCCTGCAGGCACCCCACCTGGCCCCAGAGAGAGCGTGTGG 360

Qy 361 CACGGCGTCCGACCCCCAGCGGGGCGCGCCTGACCCAGCTACTGCGAGACGTGGCCGACG 420
        ||||||||| || |||||| ||||||| ||||||| ||||||| | |||| |||||||
Db 361 CATGGGTTCGGACCCCCAACGGGCGCAGGCTGACCGAGAGCTACTGCGAGACGTGGCGGGACG 420

Qy 421 GAGGCCCCGGGCGCCACCGGGCCACAACCGGGCAGCAGCGGCTCGGCCTCGTGCTGGGAGCAG 480
        |||| || |||||||| || ||| ||||| |||| ||||| ||| ||||||| |||||||||
Db 421 GAGGCTCCCTGCCCGGCGCACGGGGCCGACGGGCCAGGGGGCAGCAGGCTCCTGGGGCGCAG 480

Qy 481 GAGGCCGCGAGCTGCCGGCCACGGCTTCGTGGTGCTCTGCATCGAGAACAGCGTCATGACC 540
        |||||||||||||||| |||| ||||||||||| |||||||| ||||||||||||||||
Db 481 AGTGCCGGCGAGCTGCCATCACGGCTACATCGCTGCTCTGCATGAGAACAGCTTCATGACT 540

Qy 541 TCCTTC 546
        ||||
Db 541 GCCTCC 546

RESULT 9
AAS00867
ID   AAS00867 standard; DNA; 549 BP.
XX
AC   AAS00867;
XX
DT   04-JUL-2001 (first entry)
XX
DE   Human gene fragment encoding Endostatin(TM) protein.
XX
KW   Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;
KW   blood borne tumour; leukaemia; tumour metastasis; benign tumour;
KW   haemangioma; acoustic neuroma; neurofibroma; trachoma; rubeosis;
KW   pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;
KW   ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW   retinopathy of prematurity; macular corneal graft rejection;
KW   neovascular glaucoma; retrolental fibroplasia; Osler-Webber Syndrome;
KW   myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW   haemophiliac joint; angiofibroma; wound granulation; ds.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..549
FT   /*tag= d
FT   /product= "Endostatin(TM)"
FT   /partial
FT   /note="None of the above CDSs have start or stop codons"
FT   CDS   1..546
FT   /*tag= c
FT   /product= "Endostatin(TM) C-terminus minus 1"
FT   /partial
FT   /note= "Variant produced during fermentation reaction of
FT   Pichia pastoris harbouring an expression plasmid
FT   containing the present sequence"
FT   CDS   1..543
FT   /*tag= b
FT   /product= "Endostatin(TM) C-terminus minus 2"
FT   /partial
FT   /note= "Variant produced during fermentation reaction of
FT   Pichia pastoris harbouring an expression plasmid
FT   containing the present sequence"
FT   CDS   1..540
FT   /*tag= a
FT   /product= "Endostatin(TM) C-terminus minus 3"
FT   /partial
FT   /note= "Variant produced during fermentation reaction of
FT   Pichia pastoris harbouring an expression plasmid
FT   containing the present sequence"
XX
PN   WO200119989-A2.
XX
PD   22-MAR-2001.
XX
PF   14-SEP-2000; 2000WO-US025166.
XX
PR   14-SEP-1999; 99US-0153698P.
XX
PA   (ENTR-) ENTREMED INC.
XX
PI   Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ;
PI   Bermejo LL, Mistry FR, Shepard SR, Schrimsher JL;
XX

DR   WPI; 2001-244802/25.
DR   P-PSDB; AAU00896, AAU00897, AAU00898, AAU00899.
XX
PT   Producing Endostatin protein for treating angiogenesis mediated diseases
PT   such as solid tumors, comprises recombinantly producing the protein using
PT   an expression system, and recovering and purifying the protein.
XX
PS   Claim 6; Page 29; 67pp; English.
XX
CC   The sequence encodes Human Endostatin(TM). The new method of the
CC   invention is useful for producing, recovering and purifying Endostatin
CC   (TM) from biological sources, such as biological fluids, tissues, cells,
CC   culture media, and fermentation media. Endostatin(TM) is useful for
CC   treating angiogenesis mediated diseases such as solid tumours, blood
CC   borne tumours, leukaemias, tumour metastases, benign tumours, e.g.
CC   haemangioma, acoustic neuromas, neurofibromas, trachomas, and pyogenic
CC   granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases,
CC   e.g., diabetic retinopathy, retinopathy of prematurity, macular
CC   degeneration, corneal graft rejection, neovascular glaucoma, colon
CC   cancer, retrolental fibroplasia, rubeosis, Osler-Webber Syndrome,
CC   myocardial angiogenesis, plaque neovascularisation, telangiectasia,
CC   haemophiliac joints, angiofibroma, and wound granulation. Endostatin(TM)
CC   is also useful for treating disease of excessive or abnormal stimulation
CC   of endothelial cells such as intestinal adhesions, atherosclerosis,
CC   scleroderma and hypertrophic scars. Higher yields of more purified, and
CC   biologically active Endostatin(TM) are obtained by the new method.
CC   Endostatin(TM) can be stored in buffers for extended periods of time, and
CC   also subjected to lyophilisation, while preserving biological activity.
CC   Centrifugation of broth from fermentation steps in production is avoided,
CC   preventing unwanted potential cellular lysis and contamination with
CC   additional proteins, pigments, enzymes and other cellular chemicals and
CC   debris
XX
SQ   Sequence 549 BP; 82 A; 196 C; 178 G; 93 T; 0 U; 0 Other;

Query Match   75.3%;   Score 418; DB 5;   Length 549;
Best Local Similarity 85.3%;   Pred. No. 8.9e-61;
Matches 466;   Conservative   0; Mismatches   80;   Indels   0; Gaps   0;

```
Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 60
         ||||| ||||||  | ||||||||||| |||| ||||||||| | ||||||||||||
Db   1 CACAGCCACCACCGCGACTTCCAGCCCGGTGCTCCACCTGGTTGCGCTCAACAGCCCCTGTCA 60

Qy  61 GCGGGCATGCGAGGCATCCGGGGAGCCGGACTTCCAGTGCTTCAGCAGCCGCGCGCCGCG 120
         ||||||||||| || ||||| ||| ||||||||||| |||||||||||||||  | ||
Db  61 GCGGGCATGCGGGGGCATCCGCGGGGCCGACTTCCAGCTTCCAGCAGCCGCGGGGCCGTG 120

Qy 121 GGGCTGGCCGGCGACCTTCCGGGCCTTCCTGTCGTCGCGCCAGGACCTTCTACAGCATC 180
         ||||||| |||||||||||||  ||||||| |||| ||||| ||||||| ||||||||
Db 121 GGGCTGGCGGGCGACCTTCCGGCCCTTCCGCCTTCCTGTCGCGCCGCCAGGACCTTGTACAGCATC 180

Qy 181 GTGCGCGCCGCACCGCACCCGGGGTGCCCGTGTCAACCTCAGGGACGGAGGTGCTCTTC 240
         |||||| || || || |||  |  |||  |||| ||||||| |||| |||||||| ||
Db 181 GTGCGCCGTGCCGACCCGCGACCGCGCAGCCGTGCCATCGTCGTCCATCGTCAACCTCAAGGACGAGCTGCTGTTT 240

Qy 241 CCCAGCTGGGAGGCCTTATTCTCGGAGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC 300
         |||||||||||| || |||| || ||| ||||||| |||||||||||||||||||| ||||
Db 241 CCCAGCTGGGAGCTCTGTTCTTCAGGGTCTGGAGGGTCCGGCTGAAGCCCGGGGCCCACGCATC 300

Qy 301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCACCCCGGCCCCGGAGAGAGCGTGTGG 360
         ||||| ||| ||||||| || |||||| |||||||||| || |||||| |||||||||||| |||
Db 301 TTCTCCTTTGACGGCAAGGACGTCCTGAGGCACCCCACCTGGCCCCAGAGAGAGGCGTGTGG 360

Qy 361 CACGGCTCCGGACCCCAGCCGCGGCGGGCGCAGGGCGCAGCGCCTGACGCCGCGGGACG 420
         || ||||| || | || ||| ||| ||| |||| | | || | |||||| |||||||||
Db 361 CATGGCTCGGAGCTCGGAGAGCTACTGTGAGACGTGGCGGACGTGGCGGACG 420

Qy 421 GAGGCCCCGCGGCCACCGGCCACCGGGCGCTCGTCGCTGGGCGGGCAGGCTGGTGCTCGAGCAG 480
         |||||| ||| | ||| ||||| |||| |||||| | ||||| || ||||| | ||||||||||
Db 421 GAGGCTCCCCTGCCCACGGCCCAGGCCCTCGTGGGGGGCCAGGCCTCGTGGGGCAG 480

Qy 481 GAGGCCGCGAGTGCTGCCGCACAGCCGCCGCCTTGTGGTCTCGCATCGAGAACAGGTCATGACC 540
         |||||| |||| | |||| || ||||| |||| | ||| || |||||||||| | |||||
Db 481 GAGGGCATGCGAGGCATCCGCGGGGCCGACTTCCAGCTTCCAGCAGCCGCGGGGCCGTG 540
```

Db 481 AGTGCCGCGAGCTGCCATCACGCCTACATCGTGCTCTGCATTGAGAACAGCTTCATGACT 540

Qy 541 TCCTTC 546
        |||| |
Db 541 GCCTCC 546

RESULT 10
ABA00774
ID   ABA00774 standard; cDNA; 549 BP.
XX
AC   ABA00774;
XX
DT   18-MAR-2003 (first entry)
XX
DE   Human endostatin coding sequence.
XX
KW   Gene; human; plasminogen; angiostatin; neovascularisation;
KW   kringle domain; cell proliferation; viral vector; replication-defective;
KW   cancer; tumour; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..549
FT   /tag= a
FT   /product= "Endostatin"
XX
PN   WO200288173-A2.
XX
PD   07-NOV-2002.
XX
PF   29-APR-2002; 2002WO-US013461.
XX
PR   30-APR-2001; 2001US-0287673P.
PR   05-APR-2002; 2002US-0370634P.
XX
PA   (CELL-) CELL GENESYS INC.
XX
PI   Chang B, Wu WW, Macarthur J, Patel S, Jooss K, Mendez M;
XX
DR   WPI; 2003-129131/12.
DR   P-PSDB; AAG79753.
XX
PT   New recombinant viral vector expressing human angiostatin useful for
PT   inhibiting angiogenesis in a mammalian subject with cancer or tumor.
XX
PS   Example 4; Page 80-82; 83pp; English.
XX
CC   This sequence encodes endostatin. Endostatin is a 20 kD C-terminal
CC   fragment of collagen XVIII that inhibits angiogenesis. The endostatin
CC   coding sequence may be used in the recombinant viral vector of the
CC   invention for obtaining angiostatin activity. The vector comprises a
CC   promoter capable of expressing human angiostatin operably linked to a
CC   structural gene encoding one or more domains of human angiostatin. The
CC   vector, which may be a replication-defective viral vector, is useful for
CC   inhibiting angiogenesis in a mammal, especially with cancer or a tumour.
CC   Note: This sequence is given incorrectly in the sequence listing of the
CC   specification as an amino acid sequence
XX
SQ   Sequence 549 BP; 82 A; 196 C; 178 G; 93 T; 0 U; 0 Other;

Query Match   75.3%;   Score 418; DB 7;   Length 549;
Best Local Similarity 85.3%;   Pred. No. 8.9e-61;
Matches 466;   Conservative   0; Mismatches   80;   Indels   0; Gaps   0;

```
Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 60
         ||||| ||||||  | ||||||||||| |||| ||||||||| | ||||||||||||
Db   1 CACAGCCACCACCGCGACTTCCAGCCCGGTGCTCCACCTGGTTGCGCTCAACAGCCCCTGTCA 60

Qy  61 GGGGCATGCGGAGGCATCCGGGGGCCGACTTCCAGCTTCCAGCAGCCGCGCGCCGCG 120
         |||||||||||  || ||||| ||||||||||||||||||||||||||  | |||
Db  61 GGGGCATGCGGTGGGGGCATCCGCGGGGCCGACTTCCAGCTTCCAGCAGCCGCGGGGCCGTG 120
```

```
CC   combinations of angiostatin, endostatin, interferon, thrombospondin,
CC   interferon-inducible protein and platelet factor 4, and have anti-
CC   angiogenic and/or anti-tumor activity. The multifunctional protein may
CC   exhibit useful properties such as having similar or greater biological
CC   activity when compared to a single factor or by having improved half-life
CC   or decreased adverse side effects, or a combination of these properties.
CC   The proteins can be used for treating an angiogenic-mediated disease,
CC   e.g. cancer, diabetic retinopathy, macular degeneration, or arthritis.
CC   They can also be used for inhibiting the production of tumor cells
CC   (characteristic of lung, breast, ovarian, prostate, pancreatic, gastric,
CC   colon, renal, bladder cancers; melanoma, hepatoma, sarcoma and lymphoma)
CC   in a patient and for inhibiting tumor growth. The present sequence is
CC   used in the course of the invention
XX
SQ   Sequence 552 BP; 83 A; 196 C; 179 G; 94 T; 0 U; 0 Other;

Query Match             75.3%;  Score 418;  DB 2;  Length 552;
Best Local Similarity   85.3%;
Matches  466;           Conservative   0;  Mismatches  80;  Indels  0;  Gaps  0;

Qy    1 CACACCCACCAGGACTTCCAGCTGGTGCTGACCGTGGTGCCCGTGAACAGCCCGCAGCCG 60
          ||||| ||||| |||||||||| |||||||  || ||||||||| |||||||||||
Db    1 CACAGCCCACCGGACTTCCAGCGGGTGCTCTCCACCGTGTTGCCTCACCAGCCCCTGTCA 60

Qy   61 GGGGCATGGCGAGGCATCCGGGGAGGCGGACTTCCAGTGCTTCCAGCAGGAGGCCGCGCCGCG 120
          |||||||||||| ||||||||||||||||||||||||||||||
Db   61 GGGGCATGGCGGGGCATCCGGGGGGCGGACTTCCAGTGCTTCCAGCAGGAGGCCGCCGTG 120

Qy  121 GGGCTGGCCGGCACCTTCCTGTCGTCGGCTGCAGGACCTCTACAGCATC 180
          ||||||| || ||||||||||||||| ||| |||||||||||||||||
Db  121 GGGCTGGCGGGGCACCTTCTGCGGCCTTCCTGCCTTGCAGGACCTCTACAGCATC 180

Qy  181 GTGCCCGGTGCCGACCGACCGGCAGCCGTGCCATCGTCAACCTCAAGGACGAGGTGCTGTTT 240
          |||||||||||||||||||||||||||||||||||| ||||||||||||
Db  181 GTGGCCCGGTGCCGACCGACCGGCAGCCGTGCCATCGTCAACCTCAAGGACGAGGTGCTGTTT 240

Qy  241 CCCAGCTGGGAGGCCTTATTCTTCGGGTGCTCCGAGGCCAGCTGAAGCCGTGAAGCCCGCATC 300
          ||||||||||||||||||||||| |||||||| |||||||||||||||||||||||||
Db  241 CCCAGCTGGGAGGCTCTGTTTCTTCAGGGCTCCGCTGAGGGTCCGCTGAAGCCCGCACCGCATC 300

Qy  301 TTCTCTTTCGACGGCAGAGATGTCCTGCTCGGGGAGGCGGGACGTACTGGGCGACACGACG 360
          |||||||||||||||||||||||  |||||||||||||||||||||||||||||
Db  301 TTCTCCTTTGACGGCAAGGACGACCCAGGCCACCCACCTGACCGACAGCGTGGGCAGCAG 360

Qy  361 CACGGCTCGACCCCAGCGGCGCCGCGCCAGCCTACTGCAGACTACTGGGCGACACGACG 420
          ||||||||||||||||||| ||||||||||||||||||||||||||
Db  361 CATGGCTCGACCCCAGCCAGGGCCAGGCTACTGTGAGACGTGTGAGACGTGGGCGACACGACG 420

Qy  421 GAGGCCCCGGCGGCCACCGGGCCAGGGCCCGGCTCCTGGGGGGGGCAGCTGCTGGAGCAG 480
          ||||||||||||||||||||||||||||||||||||||||
Db  421 GAGGCCTCGGACCCCACCGGGCAGGCCTCGCTGCTCGGTGGGCCAGGCAGGCAGCAG 480

Qy  481 AGTGCCGGCGAGCTGCCATCACGCCTACATCGTGCTCGCATTGAGAACAGCTTCATGACT 540
          |||||||||||||||||||||||||||||||||||||||||
Db  481 AGGCCGGGAGTGCGCCGCCACCGCCTTGTGGTGCTTCGCATCGAGAACAGCGTCATGACC 540

Qy  541 TCCTTC 546
          |||
Db  541 GCCTCC 546


RESULT 12
AAA68203
ID   AAA68203 standard; cDNA; 552 BP.
XX
AC   AAA68203;
XX
DT   27-OCT-2000  (first entry)
XX
DE   Human endostatin cDNA sequence.
XX
KW   Angiogenesis-inhibiting protein receptor; angiogenesis; angiostatin;
KW   endostatin; plasminogen; laminin; treatment; wound healing; solid tumour;
```

```
RESULT 11
AAX35375
ID   AAX35375 standard; DNA; 552 BP.
XX
AC   AAX35375;
XX
DT   16-JUL-1999  (first entry)
XX
DE   SEQ ID 50 of WO9916889.
XX
KW   Angiostatin; endostatin; interferon; thrombospondin;
KW   interferon-inducible protein; platelet factor 4; anti-angiogenic;
KW   anti-tumor; multifunctional protein; angiogenic-mediated disease; cancer;
KW   diabetic retinopathy; macular degeneration; arthritis;
KW   tumor cell production; ss.
XX
OS   Homo sapiens.
XX
PN   WO9916889-A1.
XX
PD   08-APR-1999.
XX
PF   30-SEP-1998;  98WO-US020464.
XX
PR   01-OCT-1997;  97US-0060609P.
XX
PA   (SEAR ) SEARLE & CO G D.
XX
PI   Bolanowski MA, Caparon MH, Casperson GF, Gregory SA, Klein BK;
PI   McKearn JP;
XX
DR   WPI; 1999-255098/21.
XX
PT   New multifunctional proteins useful for treating angiogenic-mediated
PT   diseases.
XX
PS   Disclosure; Page 85; 121pp; English.
XX
CC   The specification describes multifunctional proteins which comprise
```

```
KW   psoriasis; scleroderma; myocardial angiogenesis; Crohn's disease;
KW   cerebral collateral; arteriovenous malformation; rubeosis; cancer;
KW   diabetic retinopathy; arthritis; wound healing; peptic ulcer;
KW   Helicobacter related disease; fracture; cat scratch fever; ss.
OS   Homo sapiens.
XX
PN   WO200032631-A2.
XX
PD   08-JUN-2000.
XX
PF   06-DEC-1999;   99WO-US028897.
XX
PR   04-DEC-1998;   98US-00206059.
XX
PA   (ENTR-) ENTREMED INC.
XX
PI   Macdonald NJ,   Sim KL;
XX
DR   WPI; 2000-412290/35.
XX
PT   New angiogenesis-inhibiting protein receptors, useful in methods for
PT   treating diseases and processes that are mediated by angiogenesis, such
PT   as solid tumors, psoriasis, scleroderma and myocardial angiogenesis.
XX
PS   Disclosure; Page 86; 100pp; English.
XX
CC   This invention relates to angiogenesis-inhibiting protein receptors, and
CC   the DNA sequences encoding them. Angiogenesis is the generation of new
CC   blood vessels into a tissue, and normally occurs in wound healing, foetal
CC   and embryonal development and the formation of the corpus luteum,
CC   endometrium and placenta. Angiostatin is a protein (see AAB16150 and
CC   AAA68202) involved in angiogenesis, and has an amino acid sequence
CC   similar to that of a plasminogen fragment (see murine plasminogen
CC   AAB16490). Angiostatin has the ability to inhibit angiogenesis.
CC   Endostatin is also an angiogenesis inhibiting protein (see AAB16451 and
CC   AAA68203). Sequences AAA68242 and AAB16522 represent coding and protein
CC   sequences of human laminin. Laminin is an angiostatin binding protein,
CC   and some of the peptides of the invention share homology with regions of
CC   laminin. Peptides AAB16452-B16521 (excluding AAB16490) are the
CC   angiogenesis-inhibiting protein receptor fragments of the invention. The
CC   peptides bind either angiostatin or endostatin and can be used in methods
CC   for treating diseases and processes that are mediated by angiogenesis,
CC   such as solid tumours, psoriasis, scleroderma, myocardial angiogenesis,
CC   Crohn's disease, cerebral collaterals, arteriovenous malformations,
CC   rubeosis, diabetic retinopathy, arthritis, wound healing, peptic ulcers,
CC   Helicobacter related diseases, fractures, placentation and cat scratch
CC   fever. They are useful for the detection and prognosis of cancer. DNA
CC   sequences A628204-A628241 encode the peptides of the invention
XX
SQ   Sequence 552 BP; 83 A; 196 C; 179 G; 94 T; 0 U; 0 Other;

Query Match             75.3%;  Score 418;  DB 3;  Length 552;
Best Local Similarity   85.3%;  Pred. No. 8.8e-61;
Matches  466;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy    1 CACACCCACCACCAGGACTTCCAGCTGGTGCTGCACCGTGGTGCCCTGGACACAGCCCGGACGCCG 60
          |||| ||||||||||||||||||||||||||| |||||||||||    |||||||||||| ||
Db    1 CACAGCCACCACCAGGACTTCCAGCTGGTGCTCCACCGTGGTGCGTCAAACAGCCCCCCTGTCA 60

Qy   61 GGGGCATGCGGAGGCATCCGGGGGAGCCGACTTCCAGTGCTTCCAGCAGCGGCGGCGCCGCG 120
          |||||||||| |||||||||||| || ||||||||||||||||||||||||||||||||
Db   61 GCGGGCATGCGGGGCATCCGGGGCGGGGCCGACTTCCAGTGCTTCCAGCAGCGGCGGCCGGTG 120

Qy  121 GGGCTGGCCGGCACCTTCCGGGCCTTCCTGTCGTCGGCCGGCTGAGCGGCCAGGACCTTCTACAGCATC 180
          |||||||| ||||||||| |||||||||||| ||| ||  |||||||||||||| ||||||||
Db  121 GGGCTGGCGGGCACCTTCCGCGCCTTCCTGTCCTCGCCGCGCTGAGCGGCCAGGACCTGTACAGCATC 180

Qy  181 GTGCGGCCGCCGCGACCCGACCGGGTGCCCGTGTCAACCTCAGGGACGAGGTGCTCTTC 240
          ||||| | ||| ||||| |||| ||||| ||| |||||||| |  |||||| |||| ||
Db  181 GTGCGCCGTGCCGACCGGCCAGCCGTGCCCATCGTCAACCTCAACGAAGACGAGCTGCTGTTT 240

Qy  241 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCTCCGAGGGCCTGAAGCCCGGGGCCCCGCATC 300
```

(right column / continuation)

```
Db  241 CCCAGCTGGTGGAGGCTGTTCTGTTCTTCAGGGTCGCGCTGAAGCCCGGGGGCCACGCCATC 300
          |||||||| ||||||| |||  |||||| || ||| ||||||||||||| ||| |||||||
Qy  301 TTCTCTTTGACGGCAGAGATGTCCTGCACCACCCCGCCTGCCCCCGAAGAGCGGTGTGG 360
          |||||||||||||||  | |||||||| | |||||| ||||| |||||||||||||||
Db  301 TTCTCCTTTGACGGCAAGGACGTCCTGAGGGACCCCACCTGCCCCCAGAAGAGCGGTGTGG 360
Qy  361 CACGGCTCCGACCCCGACGCGGGCGGCCCGCCGCTGGTGGCGGCTACTGCAGCACAGCTACTGGCCGACG 420
          || ||||| ||||||||| | ||| | | |||||||||| ||||||||| ||| || |||||||||||
Db  361 CATGGCTCGGACCCCAACGGGGCCCAGCTGCAGCGGGTCGCTGCTGGTGGCGGCTACTGAGCAGTGGCCGACG 420
Qy  421 GAGGCCCGGCGGCCACCGGGCAGGCGTGTCGTCGCTGCTGGCGGCGGCAGGCTGCTGGAGCAG 480
          |||||  || |||||  ||||| ||   | | | |||||  |||| || |||||| ||||||
Db  421 GAGGCTCCCGGCCACGGGCCCAGGCCTCCTCCGCTGCTTCGTGGTGCTCGTGGGGGGCCAGGCCAG 480
Qy  481 GAGGCCGGAGGCGTCGCCGCCGCCCACGCTTCGTGGTGCTCGTGCATCGAGACAGGTCATGACC 540
          |||||| ||||| || | || ||||| || |||||||||||| |||||||||  | |||||| |
Db  481 AGTGCTGGAGGCTGCCATCACGCCTACATCGTGCTCTGCATTGAGAACACAGCTTCATGACT 540
Qy  541 TCCCTTC 546
          || |||
Db  541 GCCTCC 546

RESULT 13
AAC62023
ID   AAC62023 standard; DNA; 552 BP.
XX
AC   AAC62023;
XX
DT   06-MAR-2001 (first entry)
XX
DE   Nucleotide sequence of human endostatin encoded by plasmid pMAlcH#15.
XX
KW   Streptomyces sp. strain C5; SnpA; S. venezuelae; alpha-amylase;
KW   endostatin; cancer; tumour growth; angiogenesis; ss.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   CDS            1..552
FT                  /*tag= a
FT                  /product= "endostatin"
XX
PN   WO200060945-A1.
XX
PD   19-OCT-2000.
XX
PF   12-APR-2000;   2000WO-US009747.
XX
PR   13-APR-1999;   99US-0129084P.
XX
PA   (MERI ) MERCK & CO INC.
XX
PI   Desanti CL,   Strohl WR;
XX
DR   WPI; 2000-686970/67.
DR   P-PSDB; AAB30493.
XX
PT   Preparation of soluble recombinant endostatin involves transforming
PT   Streptomycete host with expression vector comprising nucleotide sequence
PT   encoding endostatin operably linked to linker and leader peptide.
XX
PS   Example 1; Fig 6; 57pp; English.
XX
CC   The present sequence encodes human endostatin. The protein is expressed
CC   in Streptomyces. Leader sequences of Streptomyces sp. strain C5 SnpA and
CC   S. venezuelae alpha-amylase proteins are linked to the N-terminal of
CC   endostatin. This ensures that endostatin protein is produced as a
CC   secreted, soluble protein which needs no refolding, is stable in the
CC   fermentation broth and is produced in large quantities. The method is
CC   used for preparing soluble recombinant human, murine or primate
CC   endostatin, which is useful in the treatment of cancer, inhibition of
```

CC tumour growth, inhibition of angiogenesis, isolation of receptors for
CC endostatin and for identification of anti-angiogenic compounds in assays.
CC The endostatin protein is produced as a secreted, soluble protein which
CC needs no refolding, is stable in the fermentation broth and is produced
CC in large quantities. Streptomycetes are amenable for cultivation in large
CC fermentations allowing for large quantities of soluble endostatin to be
CC produced
XX
SQ Sequence 552 BP; 83 A; 196 C; 179 G; 94 T; 0 U; 0 Other;

Query Match   75.3%;   Score 418; DB 3;   Length 552;
Best Local Similarity   85.3%;   Pred. No. 8.8e-61;
Matches 466; Conservative   0; Mismatches   80; Indels   0; Gaps   0;

Qy   1 CACCCCACCAGGACTTCCAGTGGTGCTGCACCTGGTGGCCCTGAACAGCCCGCAGCCG   60
           |||||||| || |||||||||||||||||||||||||| ||||||||||||||
Db   1 CACAGCCACCCGCGACTTCCAGCCGGTGCTCCAACCTGGTTGCGCTCAACAGCCCCCTGTCA   60

Qy   61 GGCGGCATGCGAGGCATCCGGGAGCCGGACTTCCAGTGCTTCCAGCAGGCGCGCGCCGCG   120
           ||||||||||| |||||||||||| ||||||||||||||||| ||||||||||
Db   61 GGGCGCATGCGGGGGCATCCGGCGGGCCACTTCCAGTGCTTCCAGCAGGCGCGCGGGCCGTG   120

Qy   121 GTGCGGCCGGCACCTTCCTGTCGTCGGCGTTGCAGGAGGTGCTGCGCGGCTTCTACAGCATC   180
           |||||||||| ||||||||||||||||||||||||||| ||||||||||||||||
Db   121 GGGCTGGGCGGGCACCTTCCGCGCCTCCGTGTCTCCGCGCGGGCCTGCAGGACCTGTACAGCATC   180

Qy   181 GTGCCGGCCGACCGGACGCCGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCCTTC   240
           |||||||||||||||||||||| |||||||||||||| ||||||||||||||||
Db   181 GTGGGCCGTGCCGACCGGCCAGCCGTCCCATCGTCAACCTCAAGGACGAGCTGCTGTTT   240

Qy   241 CCCAGCTGGGAGGCCTTATTCTCCGGGCTCCGAGGGCCCAGCTGAAGCCCGGGGCCCCGCATC   300
           ||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   241 CCCAGCTGGGAGGGTCGCTGGAGGGTCCGAGGGCCCCAGAAGAGCGTGTGG   300

Qy   301 TTCTCCTTTCGACGGCAGAGATGTCCTCGACCACCCCGGCCCTGGCCCCGGAAGAGCGTGTGG   360
           |||||||||| |||| |||||| |||||||||||||||||||||||||||||||
Db   301 TTCTCCTTTGACGGCAAGGACGTCCTGAGGGTCGCTGAGGGTCCCGCCCTGGCCCGCGGACGC   360

Qy   361 CACGGCTCCGACCCCAGCGGCGGCCGCCCGCTGACAGCTACTGCGAGACGTGGCGGACG   420
           ||||||||||||||||||| |||||||| ||||||||||||| ||||||||
Db   361 CATGGCTCGGACCCCAGCGGCCAGGCGGCCCAGCGGCCACTGGTGGCGGCAGGCTCTGGAGCAG   420

Qy   421 GAGGCCCGACGCCGGCCACCGGCCACCGGCCTCGGCGGGCAGGCTCCTGGGGCAG   480
           |||||||||||||||||||||||||||||||| |||||||||||||||
Db   421 GAGGCTCCCTGCAGCGGCAAGGACGTCCTGAGGGGCACCCCACCTGCGCTCGGCGTGCTGGGGCAG   480

Qy   481 GAGGCCGCCAGCGTGCGCGCACCCTTCGTGGTTGCTCTGCATCGAGAACAGCGTCATGACC   540
           ||||||| |||||||||||||||||||||| ||||||| ||||||||
Db   481 AGTGCCGCGAGCTGCCATCAGCCTCGTGTCTCGCATTGAGAACAGCTTCATGACT   540

Qy   541 TCCTTC 546
           ||||
Db   541 GCCTCC 546

RESULT 14
AAC8289
ID AAC8289 standard; DNA; 552 BP.
XX
AC AAC8289;
XX
DT 02-MAR-2001 (first entry)
XX
DE Human endostatin coding sequence SEQ ID NO: 1.
XX
KW Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken; cancer;
KW inflammation; angiogenesis-dependent disease; ds.
XX
OS Homo sapiens.
XX
PN WO20006771-A1.
XX
PD 16-NOV-2000.

XX
PP 02-MAY-2000; 2000WO-US012063.
XX
PR 06-MAY-1999; 99US-0129907P.
PR 14-JUL-1999; 99US-00353333.
XX
PA (BURN-) BURNHAM INST.
XX
PI Vuori K;
XX
DR WPI; 2001-040937/05.
DR P-PSDB; AAB49379.
XX
PT Endostatin peptide comprising at least four endostatin amino acid
PT residues are e.g. angiogenesis inhibitors for treating cancer and
PT diabetic retinopathy.
XX
PS Disclosure; Page 111-112; 146pp; English.
XX
CC The present invention provides endostatin peptides which can be used in
CC the modulation of angiogenesis. This is useful in the treatment of
CC cancers, inflammation, rheumatoid arthritis, chronic articular
CC rheumatism, psoriasis, disorders associated with inopportune invasion of
CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy
CC of prematurity, macular degeneration, corneal graft rejection,
CC retrolental fibroplasia, rubeosis, capillary proliferation in
CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent
CC diseases include Osler-Webber syndrome, myocardial angiogenesis, plaque
CC neovasculisation, telangiectasia, haemophiliac joints and wound
CC granulation. In addition, the peptides can be used as birth control
CC agents
XX
SQ Sequence 552 BP; 83 A; 196 C; 179 G; 94 T; 0 U; 0 Other;

Query Match   75.3%;   Score 418; DB 5;   Length 552;
Best Local Similarity   85.3%;   Pred. No. 8.8e-61;
Matches 466; Conservative   0; Mismatches   80; Indels   0; Gaps   0;

Qy   1 CACACCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCCCTGAACAGCCCGCAGCCG   60
           |||||||| || |||||||||||||||||||||||||| |||||||||||||
Db   1 CACAGCCACCCGCGACTTCCAGCCGGTGCTCCAACCTGGTTGCGCTCAACAGCCCCCTGTCA   60

Qy   61 GGCGGCATGCGGAGGCATCCGGGAGCCGGACTTCCAGTGCTTCCAGCAGGCGCGCGCCGCG   120
           ||||||||||| ||||||||||||| |||||||||||||||| ||||||||||
Db   61 GGGCGCATGCGGGGGCATCCGGCGGGCCGACTTCCAGTGCTTCCAGCAGGCGCGCGGGCCGTG   120

Qy   121 GTGCGGCCGGCACCTTCCTGTCGTCGGCGTTGCAGGAGGTGCTGCGCGGCTTCTACAGCATC   180
           |||||||||| ||||||||||||||||||||||||||| ||||||||||||||||
Db   121 GGGCTGGGCGGGCACCTTCCGCGCCTCCGTGTCTCCGCGCGGGCCTGCAGGACCTGTACAGCATC   180

Qy   181 GTGCGGCCGGCACCGACCGGGGTGCCCGTCGTCGTCAACCTCAGGGACGAGGTGCTCCTTC   240
           ||||||||||||||||||||||||| |||||||||||||| |||||||||||||
Db   181 GTGGGCCGTGCCGACCGGCCAGCCGTGCCATCGTCAACCTCAAGGACGAGCTGCTGTTT   240

Qy   241 CCCAGCTGGGAGGCCTTATTCTCCGGGCTCCGAGGGCCCAGCTGAAGCCCGGGGCCCCGCATC   300
           |||||||||||||||||||||||| |||||||||||||||||||||||||
Db   241 CCCAGCTGGGAGGGTCGTTTCCTGAGGGTCGCTGAGGGTCGCTGAGGGCCCCAGAAGAGCGTGTT   300

Qy   301 TTCTCTTTCGACGGCAGAGATGTCCTCGACCACCCCGGCCCTGGCCCCGGAGAGCGTGGCGGACG   360
           |||||||| |||| |||||| ||||||||||||||||| ||||||||||||||||
Db   301 TTCTCCTTTGACGGCAAGGACGTCCTGAGGGTCGCTGAGGGTCCCGCCCTGGCCCGCGGACGC   360

Qy   361 TTCTCTTTCGACGGCAGAGATGTCCTCGAGCACGCCCGGCCCTGGCGGGGACGGTGTGG   420
           ||||||||| ||||| |||||||||||||||| |||||||||||||||||||
Db   361 TTCTCCTTTGACGGCAAGGACGTCCTGAGGGTCGCTGAGGGTCGCTGAGAAGAGCGGTGTGG   420

Qy   421 GAGGCCCGGCGCGCCACCGGGCACCGGCCTCGCTGCTGGGGCAGGCAG   480
           ||||||| ||||||||||||||||||||||||| |||||||||||
Db   421 GAGGCTCCCTGCAGCGGCAAGGACGTCCTGAGGGGCACCCCACCTGCGCTCGGCGTGCTGGGGCAG   480

Qy   481 GAGGCCGCCAGCGTGCGCGCACCCTTCGTGGTTGCTCTGCATCGAGAACAGCGTCATGACC   540
           ||||||| ||||||||||||||||||||| |||||||| ||||||||
Db   481 AGTGCCGCGAGCTGCCATCACGCTACATCGTGCTCTGCATTGAGAACAGCTTCATGACT   540

```
Qy      541 TCCTTC 546
            ||| |
Db      541 GCCTCC 546

RESULT 15
ABK50685
ID   ABK50685 standard; cDNA; 552 BP.
XX
AC   ABK50685;
XX
DT   13-AUG-2002  (first entry)
XX
DE   cDNA encoding human endostatin.
XX
KW   Human; angiogenesis; PITSLRE protein kinase; cancer; arthritis;
KW   macular degeneration; diabetic retinopathy; angiogenic-related disease;
KW   haemangioma; blood borne tumour; leukaemia; neovascularisation;
KW   coronary collateral; cerebral collateral; neovascular glaucoma;
KW   corneal disease; wound healing; Helicobacter related disease; fracture;
KW   keloid; ovulation; menstruation; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..552
FT                   /*tag= a
FT                   /partial
FT                   /product= "Endostatin"
FT                   /note= "This sequence lacks a start codon"
XX
PN   WO200230982-A2.
XX
PD   18-APR-2002.
XX
PF   15-OCT-2001; 2001WO-US032437.
XX
PR   13-OCT-2000; 2000US-0240127P.
XX
PA   (ENTR-) ENTREMED INC.
XX
PI   Sim KL, Liang H;
XX
DR   WPI; 2002-435440/46.
DR   P-PSDB; AAU97132.
XX
PT   Regulating angiogenesis for treating scleroderma, leukemia, keloids by
PT   administering a protein that is homologous to PITSLRE protein kinase and
PT   an angiogenic factor or a protein kinase and its active fragments.
XX
PS   Disclosure; Fig 2B; 45pp; English.
XX
CC   The present invention relates to methods and compositions of inhibiting
CC   angiogenesis. The method comprises administering to a human or animal a
CC   composition comprising a protein that is homologous to PITSLRE protein
CC   kinases (PK) and an angiogenic factor. The method is useful for
CC   regulating angiogenesis related to cancer, arthritis, macular
CC   degeneration, and diabetic retinopathy. The compositions are useful for
CC   inhibiting angiogenic-related diseases. The method and compositions are
CC   useful in treating diseases and processes that are mediated by
CC   angiogenesis including haemangioma, solid tumours, blood borne tumours,
CC   leukaemia, metastasis, telangiectasia, psoriasis, scleroderma, pyogenic
CC   granuloma, myocardial angiogenesis, Crohn's disease, plaque
CC   neovascularisation, coronary collaterals, cerebral collaterals,
CC   arteriovenous malformations, ischaemic limb angiogenesis, corneal
CC   diseases, rubeosis, neovascular glaucoma, diabetic retinopathy,
CC   retrolental fibroplasia, arthritis, diabetic neovascularisation, macular
CC   degeneration, wound healing, peptic ulcer, Helicobacter related diseases,
CC   fractures, keloids, vasculogenesis, haematopoiesis, ovulation,
CC   menstruation, placentation, and cat scratch fever. The method of the
CC   invention provides a therapy for cancer that has minimal side effects.
CC   The present sequence encodes human endostatin which is used to generate
```

```
CC   angiogenesis-inhibiting peptides
XX
SQ   Sequence 552 BP; 83 A; 196 C; 179 G; 94 T; 0 U; 0 Other;

     Query Match          75.3%;   Score 418;   DB 6;   Length 552;
     Best Local Similarity  85.3%;   Pred. No. 8.8e-61;
     Matches  466;   Conservative   0;   Mismatches   80;   Indels   0;   Gaps   0;

Qy     1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCCCTGAACAGCCCGCAGCCG 60
           ||||||||| |||||||||||||||||||||||||||| |||||||| |||||| |||| |
Db     1 CACAGCCACCGCGACTTCCAGCCGCGTGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA 60

Qy    61 GGCGGCATGCGAGGCATCCGGGGAGACGGACTTCCAGTGCTTCCAGCAGGCGCGCGCCGCG 120
           ||||||||||||| ||||||||||||| ||||||||||||||||||| ||||||||| |||
Db    61 GGCGGCATGCGGGGCATCCGCGGGGCCGACTTCCAGTGCTTCCAGCAGGCGGCGGCCGGTG 120

Qy   121 GGGCTGCCGGGACCTTCCGCGGCTGTCGTCGGCGCTGCAGGAGGACCTCTACAGCATC 180
           ||||||||||||||||||| |||||||||||||||| |||| |||||||||||||||||
Db   121 GGGCTGGCGGGCACCTTCCGCGCCTTCCTGTCCTCCGCGGACCTGCAGGACCTGTACAGCATC 180

Qy   181 GTGCGGCCGGCGACCGCACCGGGGTGCCCGTGCTCAACCTCAGGACGACGAGGTGCTCTTC 240
           |||||||||||||||||||||||||||||||||||||||||||| |||||||||| |||||
Db   181 GTGCGGCCGGTGCCGACCGCACCGGGGTGCCCGTGCTCAACCTCAAGGACGAGCTGCTGTTT 240

Qy   241 CCCAGTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC 300
           |||||||| ||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   241 CCCAGCTGGGAGGGCTCTGTTCTCAGGCTCCGAGGGTCCCGCGAGGGTCCCCGGAAGAGCCGCACGACG 300

Qy   301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCGCCTGGCCCCGAAGAGAGCGTGTGGG 360
           ||||| ||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db   301 TTCTCCTTGACGGCAAGAGACGTCCTGAGGCACCCCACCTGGCCCCACGGGCCGCGTGTGGG 360

Qy   361 CACGGCTCCGACACCCAGCTGCGGCGCGCCGCCTGACCGACAGCTACTGCGAGACGTGGCCGGACG 420
           || |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   361 CATGGCTCGGACCCCCAACCGGCCGCAGGCTGACCGAGAGCTACTGTGAGACGTGCGGGACG 420

Qy   421 GAGGCCCCGGCGGCCACCGGACCGGCAGGCGTCGTGCGGGGCAGGCTGCTGCTGGGAGCAG 480
           ||||| ||| ||||||||||||||||||||| |||||||||| |||||||| |||| ||||
Db   421 GAGGCTTCCCTCGGCCACCGGCGGCCCTGGCAGCTCCTGCTGTCTGGGGGGCCAGGCTCCTGGGGCAG 480

Qy   481 GAGGCCGCGAGCTGCCGCCCACGCCTTCGTGGTGCTCTCGCATCGAGAACAGCGTCATGACC 540
           |||||||||||||||||||| ||||||||||| || ||||| |||||| ||||||||||||
Db   481 AGTGCCGCGAGCTGCCATCACGCCTACACTGTGTCTCGCATTGAGAACAGCAGCTTCATGACT 540

Qy   541 TCCTTC 546
           ||| |
Db   541 GCCTCC 546
```

```
Search completed: March 29, 2004, 12:05:45
Job time : 249.424 secs
```

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

Run on: March 29, 2004, 10:57:20 ; Search time 56.1416 Seconds
                                  (without alignments)
                                  5486.086 Million cell updates/sec

Title: US-09-938-391-3
Perfect score: 555
Sequence: 1 cacacccaccaggacttcca..........tgacctccttcccaagtag 555

Scoring table: IDENTITY.NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
    1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
    2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
    3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
    4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
    5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
    6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	418	75.3	546	4	US-09-315-689-4	Sequence 4, Appli
2	418	75.3	552	3	US-09-206-059-30	Sequence 30, Appli
3	418	75.3	563	4	US-09-231-077D-9	Sequence 9, Appli
4	418	75.3	3394	1	US-08-159-784-4	Sequence 4, Appli
5	410.8	74.0	534	3	US-09-315-689-6	Sequence 6, Appli
6	375.8	67.7	558	3	US-09-449-293-3	Sequence 3, Appli
7	375.8	67.7	558	4	US-09-775-325-3	Sequence 3, Appli
8	375.8	67.7	558	4	US-09-775-174-3	Sequence 3, Appli
9	375.8	67.7	565	3	US-08-998-526-37	Sequence 37, Appli
10	374.2	67.4	4031	1	US-08-159-784-1	Sequence 1, Appli
11	373.4	67.3	573	4	US-09-561-500-12	Sequence 12, Appli
12	373.4	67.3	573	4	US-09-561-108-12	Sequence 12, Appli
13	373.4	67.3	573	4	US-09-561-526-12	Sequence 12, Appli
14	373.4	67.3	573	4	US-09-561-499-12	Sequence 12, Appli
15	373.4	67.3	573	4	US-09-998-831-12	Sequence 12, Appli
16	373.2	67.2	568	4	US-09-231-077D-8	Sequence 8, Appli
17	373.2	67.2	580	4	US-09-231-077D-7	Sequence 7, Appli
18	371.6	67.0	582	4	US-09-231-077D-6	Sequence 6, Appli
19	370.8	66.8	620	4	US-09-231-077D-5	Sequence 5, Appli
20	60.8	11.0	1116	4	US-09-252-991A-2102	Sequence 2102, Ap
21	60.8	11.0	1224	4	US-09-255-991A-1793	Sequence 1793, Ap
22	59.4	10.7	1200	4	US-09-252-991A-1861	Sequence 1861, Ap
23	59.4	10.7	1722	4	US-09-252-991A-2173	Sequence 2173, Ap
24	57.4	10.3	741	4	US-09-252-991A-6774	Sequence 6774, Ap
25	57.4	10.3	1557	4	US-09-252-991A-6920	Sequence 6920, Ap
26	57.4	10.3	1590	4	US-09-252-991A-6824	Sequence 6824, Ap
27	55.8	10.1	1820	1	US-08-173-508-7	Sequence 7, Appli
28	55.8	10.1	1821	2	US-08-265-310-7	Sequence 7, Appli
29	55.8	10.1	1821	3	US-08-951-742-7	Sequence 7, Appli
30	53.8	9.7	1050	3	US-09-252-991A-10973	Sequence 10973, A
31	53.8	9.7	2547	4	US-09-252-991A-11271	Sequence 11271, A
32	53.8	9.7	2964	4	US-09-252-991A-10912	Sequence 10912, A
33	53.6	9.7	1140	4	US-09-522-714-13	Sequence 13, Appli
34	53	9.5	711	4	US-09-252-991A-11047	Sequence 11047, A
35	53	9.5	15872	3	US-09-105-537-1	Sequence 1, Appli
36	53	9.5	15872	4	US-09-091-609-1	Sequence 1, Appli
37	53	9.5	15872	4	US-09-091-609-3	Sequence 3, Appli
38	52.4	9.4	30001	1	US-08-125-468-1	Sequence 1, Appli
39	52.4	9.4	30001	2	US-08-474-933-1	Sequence 1, Appli
40	52	9.4	20235	3	US-07-642-734C-3	Sequence 3, Appli
41	52	9.4	20235	3	US-08-439-009A-3	Sequence 3, Appli
42	51.8	9.3	1260	4	US-09-252-991A-8957	Sequence 8957, Ap
43	51.8	9.3	1314	4	US-09-252-991A-9048	Sequence 9048, Ap
44	51.8	9.3	2133	3	US-09-252-991A-9129	Sequence 9129, Ap
45	49.4	8.9	71989	4	US-09-443-501A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-315-689-4
; Sequence 4, Application US/0915689
; Patent No. 6345510
; GENERAL INFORMATION:
; APPLICANT: Folkman, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
; FILE REFERENCE: 05213-0229
; CURRENT APPLICATION NUMBER: US/09/315,689
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 4
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-315-689-4

Query Match   75.3%; Score 418; DB 4; Length 546;
Best Local Similarity   85.3%; Pred. No. 1.3e-74;
Matches 466; Conservative   0; Mismatches 80; Indels   0; Gaps   0;

```
Qy    1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCTGAACAGCCCGACAGCCG  60
          |||||| ||||| ||||||||||||| || ||||||||||||||||||||||||||||
Db    1 CACAGCCACCGGACTTCCAGCCGTGCTCCACCTGGTTGCGTCAACAGCCCCTGTCA  60

Qy   61 GGCGGCATGGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGGCGCGCGGCGCG  120
          ||||||||||||||||||||||||| ||||||||||||| ||||||||||||||||
Db   61 GGCGGCATGGAGGGCATCCGCGGGGGCCGACTTCCAGTGCTTCCAGCAGGCGGGGCGTG  120

Qy  121 GGGCTGGCGGCACCTTCCGTCTCGTCGGCTGCGAGGACCTCTACAGCATC  180
          |||||| |||||||||||| |||||||||||| ||||||| ||||||||
Db  121 GGGCTGGCGGGGCACCTTCCGGCCTTCCGTCCTCGGCCGGACCTGTACAGCATC  180

Qy  181 GTGCGCCGCGGCCGCACGGCACCGGGGTGCCCGTGCTGCAACCTCAGGGACGAGGTGCTTC  240
          ||||||| ||||| ||||| ||||||||||| |||| ||||||||||||| ||||||
Db  181 GTGCGCCGTGCCGCGACCGGCCCAGCCGTGCCCATCGTCAACCTCAAGGACGAGCTGCTGTTT  240

Qy  241 CCCAGCTGGGAGGCCTTATTCTCGGCGTCCGAGGGCCAGCTGAAGCCCGGGCCCGCATC  300
          |||||||||||||||| || |||||||| |||||||||||||||||||||||||||
Db  241 CCCAGCTGGGAGGCTCGTTCTTCAGGCTCTGAGGGTCCGCTGAAGCCCGGGCACGCATC  300

Qy  301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCGCTGCGAGCAC GTGTGG  360
          ||||||| ||||||||||| || |||||||||||||||||||||||||||
Db  301 TTCTCCTTTGACGGCAAGGACGTGCCTGAGGTCGCACCCCACCTGGCGCCAGAAGAGCGTGTGG  360

Qy  361 CACGGCTCCGACCCCAGCGGCGGCCGCTACTGCGAGGACGTGGCGGACG  420
          |||||||||||||||||||||||| ||||||||||||||||||||||||
Db  361 CATGCGCTCGGACCCCAACGGCGGCCAGCGCCAGCGCTACTGTGAGACGTGGCGGACG  420
```

Qy   541 TCCTTC 546
           ||||||
Db   541 GCCTCC 546

RESULT 2
US-09-206-059-30
; Sequence 30, Application US/09206059
; Patent No. 620104
; GENERAL INFORMATION:
;   APPLICANT: MacDonald, Nicholas
;   APPLICANT: Sim, Kim Lee
;   TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and
;   TITLE OF INVENTION: Proteins and Methods of Use
;   FILE REFERENCE: 05213-0370
;   CURRENT APPLICATION NUMBER: US/09/206,059
;   CURRENT FILING DATE: 1998-12-04
;   NUMBER OF SEQ ID NOS: 80
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
;   LENGTH: 552
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-206-059-30

Query Match   75.3%; Score 418; DB 3; Length 552;
Best Local Similarity   85.3%; Pred. No. 1.3e-74;
Matches 466; Conservative   0; Mismatches 80; Indels 0; Gaps 0;

Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCTGGCCCTGAACAGCCCGCAGCCG 60
          ||||||| |||||| ||||||||| ||||||| ||||||||||||||||||||| |||||||
Db   1 CACAGCCACCGCGGGCACTTCCAGCCGGTGCTCCACCTGGTTGCGGCTCCAACAGCCCCCGTCA 60

Qy   61 GGCGGCATGCGAGGGCATCCGGGGGGCGACTTCCAGTGCTTCCAGCGAGGCGCGGCCGTG 120
          |||||| |||||| ||||||||| |||||||||| |||||||||||| ||||||||||||
Db   61 GGCGGCATGCGGGGGCATCCGGGGCCGACTTCCAGTGCTTCCAGCGAGGCGGCGGCCGTG 120

Qy   121 GGGCTGGCCCGCACCTTCCGGGCCTTCCACGTGCTGGCCTGCGCGCTGCGCTCTACAGCATC 180
          |||||||||||| |||||||||| ||||||||||||| |||||||||||||||||||||||
Db   121 GGGCTGGGCGGGCACCTTCGCGCCTTCGCGGTGCTGGCCTGCGCGCTGCGCTGTACAGCATC 180

Qy   181 GTGCGCGCCGCCGGGGTGCCCGTGCCGCCCATGGCGGGCCCGGGGAAGAGCGTGTGG 240
          |||||||||||||||||||||||||||| ||||||| |||||| ||||||||||||
Db   181 GTGCGCCGTGCCGACGGCGCCGCGCCATGGCGGGCCCGGGGAAGAGCGTGTGG 240

Qy   241 CCCAGCTGGGAGGCCTTATTCTCGGAGGCTCGAGGCCAGCTGAAGCCCCGGAGCCCCGCATC 300
          |||||||||||||||||||||||||||| |||||| ||||||||||||||||||||||||
Db   241 CCCAGCTGGGAGGCTCTGTTCTCAGGCTCTGAGGGTGCCGCACCGCGGGGCACGCATC 300

Qy   301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGGCTCCCGCGGAAGAGCGTGTGG 360
          |||||||||||||||||||| |||||||||||||||||| |||||||||||||||||||
Db   301 TTCTCCTTTGACGGCAGAAGGACGTCCTGAGGCGCAGCACCCCAGCCCCCAGAGAGCGTGTGG 360

Qy   361 CACGGCTCCGACCCCAGCGGGCCGCGACCAGCTACTCCGAGACGTGCGCGACG 420
          |||||||| |||||||||||||| |||||||||||| ||||||||||||||||
Db   361 CATGGCTCCGACACCCCAACGGGCCGACCAGCTACTGTGAGACGTGCGCGACG 420

Qy   421 GAGGCCCCGGCGGCTCCCTGCCTCCTGCGTGCTTCGGGGGGGCAGGGCTCCTGGGGCAG 480
          ||||||||||| |||||||||| ||||||||||||| |||||||| |||||||||||||
Db   421 GAGGCTCCCGCCAGCGGCCCCACGGGCCACCGGAGGCTCCTGGAGCAG 480

Qy   481 GAGGCCGCGAGCTGCCATCACGCCTACACGGCGCTGCTCTGCATTGAGAAACAGCTTCATGACT 540
          |||||||||||||||||| |||||||||||||||| |||||||||||||||||||||||||
Db   481 AGTGCCGCGAGCTGCCATCACGCCTACACGGCTTCTGCATTGAGAAACAGCTTCATGACT 540

RESULT 3
US-09-231-077D-9
; Sequence 9, Application US/09231077D
; Patent No. 6653098
; GENERAL INFORMATION:
;   APPLICANT: Harding, E.I.
;   APPLICANT: Violand, B.N.
;   TITLE OF INVENTION: Method of producing mouse and human
;   TITLE OF INVENTION: endostatin
;   FILE REFERENCE: S03071-00-US
;   CURRENT APPLICATION NUMBER: US/09/231,077D
;   CURRENT FILING DATE: 1999-01-14
;   PRIOR APPLICATION NUMBER: 60/075,587
;   PRIOR FILING DATE: 1998-02-23
;   NUMBER OF SEQ ID NOS: 13
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 563
;   TYPE: DNA
;   ORGANISM: human
US-09-231-077D-9

Query Match   75.3%; Score 418; DB 4; Length 563;
Best Local Similarity   85.3%; Pred. No. 1.3e-74;
Matches 466; Conservative   0; Mismatches 80; Indels 0; Gaps 0;

Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCTGGCCCTGAACAGCCCGCAGCCG 60
          ||||||| |||||| ||||||||| ||||||| ||||||||||||||||||||| |||||||
Db   6 CACAGCCACCGCGGGCACTTCCAGCCGGTGCTCCACCTGGTTGCGGCTCCAACAGCCCCCGTCA 65

Qy   61 GGCGGCATGCGAGGGCATCCGGGGGGCGACTTCCAGTGCTTCCAGCGAGGCGCGGCCGTG 120
          |||||| |||||| ||||||||| |||||||||| |||||||||||| ||||||||||||
Db   66 GGCGGCATGCGGGGGCATCCGGGGCCGACTTCCAGTGCTTCCAGCGAGGCGGCGGCCGTG 125

Qy   121 GGGCTGGCCCGCACCTTCCGGGCCTTCCACGTGCTGGCCTGCGCGCTGCGCTCTACAGCATC 180
          |||||||||||| |||||||||| ||||||||||||| |||||||||||||||||||||||
Db   126 GGGCTGGGCGGGCACCTTCGCGCCTTCGCGGTGCTGGCCTGCGCGCTGCGCTGTACAGCATC 185

Qy   181 GTGCGCGCCGCCGGGGTGCCCGTGCCGCCCATGGCGGGCCCGGGGAAGAGCGTGTGG 240
          |||||||||||||||||||||||||||| ||||||| |||||| ||||||||||||
Db   186 GTGCGCCGTGCCGACGGCGCCGCGCCATGGCGGGCCCGGGGAAGAGCGTGTGG 245

Qy   241 CCCAGCTGGGAGGCCTTATTCTCGGAGGCTCGAGGCCAGCTGAAGCCCCGGAGCCCCGCATC 300
          |||||||||||||||||||||||||||| |||||| ||||||||||||||||||||||||
Db   246 CCCAGCTGGGAGGCTCTGTTCTCAGGCTCTGAGGGTGCCGCACCGCGGGGCACGCATC 305

Qy   301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGGCTCCCGCGGAAGAGCGTGTGG 360
          |||||||||||||||||||| |||||||||||||||||| |||||||||||||||||||
Db   306 TTCTCCTTTGACGGCAGAAGGACGTCCTGAGGCGCAGCACCCCAGCCCCCAGAGAGCGTGTGG 365

Qy   361 CACGGCTCCGACCCCAGCGGGCCGCGACCAGCTACTCCGAGACGTGCGCGACG 420
          |||||||| |||||||||||||| |||||||||||| ||||||||||||||||
Db   366 CATGGCTCCGACACCCCAACGGGCCGACCAGCTACTGTGAGACGTGCGCGACG 425

Qy   421 GAGGCCCCGGCGGCTCCCTGCCTCCTGCGTGCTTCGGGGGGGCAGGGCTCCTGGGGCAG 480
          ||||||||||| |||||||||| ||||||||||||| |||||||| |||||||||||||
Db   426 GAGGCTCCCGCCAGCGGCCCCACGGGCCACCGGAGGCTCCTGGAGCAG 485

Qy   481 GAGGCCGCGAGCTGCCATCACGCCTACACGGCGCTGCTCTGCATTGAGAAACAGCTTCATGACC 540
          |||||||||||||||||| |||||||||||||||| |||||||||||||||||||||||||
Db   486 AGTGCCGCGAGCTGCCATCACGCCTACACGGCTTCTGCATTGAGAAACAGCTTCATGACT 545

Qy   541 TCCTTC 546
          ||||||
Db   546 GCCTCC 551

RESULT 4
US-08-159-784-4
; Sequence 4, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3394
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-159-784-4

Query Match   75.3%;   Score 418;   DB 1;   Length 3394;
Best Local Similarity 85.3%; Pred. No. 1.5e-74;
Matches 466;   Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy   1 CACACCCACCAGGACTTCCAGCTGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG   60
          ||| |||| |||||| ||||| ||| ||||| |||| || ||
Db 1504 CACAGCCACCCGACTTCCAGCCGTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA 1563

Qy   61 GGCGGCATGCGAGGCATCCGGGACTTCCGGGGCCATCCCGACTTCCAGTGCTTCCAGCAGGCGCGCCG 120
          ||||||||||||| ||||| ||||||||||||
Db 1564 GGCGGCATGCGAGGGCATCCCGGGCCATCCCGACTTCCAGTGCTTCCAGCAGGCGCGGGCCGTG 1623

Qy   121 GGGCTGGCCGGCACACCTTCCGGGCCTTCTGTCGTGCGCGGGCTGCGAGGACCTCCACAGCATC 180
          ||| ||||| |||||| |||||| ||||
Db 1624 GGGCTGGCGGGCACCTTCCGCGCCTTCCGTGTCCTCGCGGACCTGTACAGCATC 1683

Qy   181 GTGCGCGCCGCGCCGACGGACACCGGGGTGCCCGTGTCAACCTCAGCGGACGAGGTGCCTCTTC 240
          ||| |||| ||||||
Db 1684 GTGCGCGTGCCGACCGCGCAGCCGTGCCCGTGCCCATGTCAACCTCAAGCGACGACGTACTGTGTTT 1743

Qy   241 CCCAGCTGGGAGGCGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGCGCCCGCATC 300
          ||||||||||||| |||||| |||||||
Db 1744 CCCAGCTGGGGAGGCTCTGTTCTCAGGCTCTGAGGGTCCGCTGAAGCCCGGGCCACGGCATC 1803

Qy   301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGGCTGGCCCCGCCCGGCGTGTGG 360
          |||||| ||| ||||||| ||||||||||
Db 1804 TTCTCCTTTGACGGCAAGGACGTCCTGAGGCACCCCACCTGGCCCCAGCAGGACGAGGCGTGTGG 1863

Qy   361 CACGGCTCCGACCCCAGCTGGGAGGCGCCGGCTACTGCGACGCAGCTACTGCGAGACGTGGCGGACG 420
          ||| ||||| |||||| |||||||||
Db 1864 CATGGCTCCGACCCCACGGGCGCAGGCTGACCGAGAGCTACTGTGAGACGTGGCGGACG 1923

Qy   421 GAGGCCCCGCGGCACCGGCAGGCGGCTCGTCGTCGTCGTCGGGCAGGCTGCTGGAGCAG 480
          ||||||| ||||||| |||||||||| ||||||||||||
Db 1924 GAGGCTCCCTCCGGCCACGGGCCAGGCCTCCTCCGCTCGTCGGGGGGCAGGCTCCTCGGGGCAG 1983

Qy   481 GAGGCCGGAGCTGCCGCCACGGCCTTCGTGGTCGTCTGGCTCTGGCATCGAGAACAGCGTCATGACC 540
          |||||| |||||||| || ||||| ||||||| |||||||| |||||||| || ||||||
Db 1984 AGTGCGCGAGCTGCCATCACGCCTACATCGTGCTCTGGCATTGAGAGCATTGAGAACAGCTTCATGACT 2043

Qy   541 TCCTTC 546
          ||
Db 2044 GCCTCC 2049

RESULT 5
US-09-315-689-6
; Sequence 6, Application US/09315689
; Patent No. 6346510
; GENERAL INFORMATION:
; APPLICANT: Folkman, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
; FILE REFERENCE: 05213-0229
; CURRENT APPLICATION NUMBER: US/09/315,689
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 534
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-315-689-6

Query Match   74.0%;   Score 410.8;   DB 4;   Length 534;
Best Local Similarity 85.6%; Pred. No. 3.5e-73;
Matches 457;   Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy   13 GACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCGGGCGCATGCGGA   72
          ||||||||| ||||||| |||||||| ||| ||||||||| |||||||||||||||
Db   1 GACTTCCAGCCGGTGCTTCCACCTGGTTGCGCTCAACAGCCCCCTGTCAGGCGGCATGCGG   60

Qy   73 GGGCATCCGGGAGGCGACTTCCAGTGCTTCCAGCAGGCGCGGGCCGCCGCGGGCTGGCCGGC 132
          || ||||| ||| | ||||||||||||||||||||||||| ||| |||||||||||| ||
Db   61 GGCATCCCGGGCCGGGGACTTCCAGTGCTTCCAGCAGGCGCGGGGCCGTGGGGCTGGCGGGGC 120

Qy   133 ACCTTCCGGGCCTTCTGTCGTGCGCGGGCTGCAGCACCTTACAGCATCGTGCGCGCCGCC 192
          ||||||| |||||| || ||| | ||| ||||||||| | ||||||||||||||| || ||
Db   121 ACCTTCCGCGCCTTCCGTGTCCTCGCGGCCTGCAGCACCGTGTACAGCATCGTGCGCGTGCC 180

Qy   193 GACCGACACCGGGGTGCCCGTGTCAACCTCAGCGGACGAGGTGCCTCTTCCCCAGCGTGGGAG 252
          ||||| | ||| |||||||||| ||||||||| |||||| || || |||| |||||||||| ||
Db   181 GACCGCGCAGCCGTGCCCGTGCCCATGTCAACCTCAAGCGACGACGTACTGTGTTTCCCAGCGTGGGGAG 240

Qy   253 GCCTTATTCTCCGGGCTCCGAGGGCCAGCTGAAGCCCGGCGCCCGCATCTTCTCTTTCGAC 312
          || | ||||| ||||| |||||| |||||||||||||| ||||| ||||||||| |||||
Db   241 GCTCTGTTCTCAGGCTCTGAGGGTCCGCTGAAGCCCGGGCCACGGCATCTTCTCCTTTGAC 300

Qy   313 GGCAGAGATGTCTGCAGCACCCCGGCTGGCCCCGCCCAGAGACGGCGTGTGGCACGGCTCCGAC 372
          |||| | || ||| || ||||||| |||||||| |||| ||||| ||| ||||| |||||||
Db   301 GGCAAGGACGTCCTGAGGCACCCCACCTGGCCCCAGCAGGACGAGGCGTGTGGCATGGCGTCGGAC 360

Qy   373 CCCAGCTGGGAGGCGCCGGCTACTGCGACGCAGCTACTGCGAGACGTGGCGGACGCCCCGGCG 432
          |||| | |||||| ||||||||| || |||| |||||| || || ||||| |||||||| |||
Db   361 CCCAACGGCGGCGGCCAGGCTGACCGAGAGCTACTGTGAGGCGACGGAGGCGCCCCTCG 420

Qy   433 GCCACCGGGCAGGGCGTCGTCGGCCTGGCGGGCAGGCTGCTGGAGCAGGCCGCGAGC 492
          |||||| ||| |||| |||| |||| |||||||||||| || ||||||| || |||
Db   421 GCCACGCGGGCCAGGCCGCCAGGCTCCTGGGGGGCAGGCTCCGCGGGCCAGGAGCTGCCGCGAGC 480

Qy   493 TGCCGCCACGCCTTCGTGGTGCTCTGCATTGAGACCTCCTTC 546
          |||| | || ||| || ||||||||||||||||| || |||
Db   481 TGCCATCAGCCTACATCGTGCTCTGCATTGAGAGCAGCTTCATGACTGCCTCC 534

RESULT 6
US-09-449-293-3
; Sequence 3, Application US/09449293
; Patent No. 6267954
; GENERAL INFORMATION:
; APPLICANT: Abitbol, Marc
; APPLICANT: Uteza, Yves
; APPLICANT: Menasche, Maurice
; APPLICANT: Bossard, Carine
; APPLICANT: Van Den Berghe, Loic
; APPLICANT: Bonnel, Sebastien
; APPLICANT: Prats, Herve
; APPLICANT: Honiger, Jiri
; APPLICANT: Neuner-Jehle, Martin
; TITLE OF INVENTION: INTRAOCULAR TRANSPLANTATION OF ENCAPSULATED CELLS
; FILE REFERENCE: 8076.2020US01
; CURRENT APPLICATION NUMBER: US/09/449,293
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Rattus rattus
US-09-449-293-3

Query Match   67.7%;   Score 375.8;   DB 3;   Length 558;
Best Local Similarity   79.8%;   Pred. No. 3e-66;
Matches 443;   Conservative   0;   Mismatches 112;   Indels   0;   Gaps   0;

Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCCCTGAACAGCCCGCAGCCG 60
            ||||| ||||||||||| |||||||||||||| |||||| ||||||| ||||||||||||
Db   4 CATACTCATCAGGACTTTCAGCCAGTGCTCCACCTGGTGGCACTGAACACCCCCCTGTCT 63

Qy   61 GGCGGCATGCGAGGCATCCGGGAGCGGACTTCCAGTGCTTCCAGCAGGCGCGCGCCGCG 120
            ||||||||| || |||||||||| ||||| ||||||| |||||||| ||| |||||||
Db   64 GGAGGCATGCGTGGTATCCGTGGCAGATTTCCAGTGCTTCCAGCAAGCCCGAGCCGTG 123

Qy   121 GGGCTGGCCCGCACCTTCCGGGCCTTCGTGTGTGCGGCTGCAGGACCTCTACAGCATC 180
            ||||| ||||||||||| ||||| |||||||| ||||| |||||| ||||||||||||
Db   124 GGGCTGTCGGGCACCTTCCGGGGCTTTCCTGTCCTCTAGGCTGCAGGATCTCTATAGCATC 183

Qy   181 GTGCGCGCGCCGACCGCACCGGGGTGCCGTCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 240
            |||||| ||||||||| ||||||| ||||| |||||| ||||||||||||| ||||||||
Db   184 GTGCGCCGTGCTGCACCGGGGTCTGTGCCCATCGTCAACCTGAAGGACGAGGTGCTATCT 243

Qy   241 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGCCCGCATC 300
            ||||| |||| ||||| |||||| |||||| ||| ||||| |||||||||| ||||||
Db   244 CCCAGCTGGGACTCCCTGTTTTCTCGGCTCCCAGGGTCAACTGCAACCCGGGGCCCGCATC 303

Qy   301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCTGGCCCCGGAAGAGCGTGTGG 360
            ||||| |||||||| |||||||||||||| ||| |||||| |||||||||| ||||||
Db   304 TTTTCTTTCGACGGCAGAGATGTCCTGAGACACCACGACACCCACCCAGCGCCGTATGG 363

Qy   361 CACGGCTCCACCCCAGCCGGCGCCGCCTGACCAGCTACTGCGAGACGTGGCCGGACG 420
            ||||| ||||||||||||| | |||||| || |||| ||||| ||||||||||||||
Db   364 CACGGCTCGGACCCCAGTGGGCGGAGGCTGATGGAGCATGCTCGTGCTGTGGAGACT 423

Qy   421 GAGGCCCCGGCGGCCACCGGGGCAGGCGTCGTGCTGGCGGGCAGGCTGCTGGAGCAG 480
            |||||| |||| ||| ||||||| | ||||||||||| |||||||| ||||||||||
Db   424 GAAACTACTGGGGCTACAGTGCTCAGGCCTCTCCCTGCGTGTCCTGGAACAG 483

Qy   481 GAGGCCGCGAGCTCCCGCCACCGCCTTCGTGGTGCTCTGCATCGAGAACAGCGTCATGACC 540
            |||||||| ||||||||| |||||||| ||||||||||| ||||| |||||||||||||
Db   484 AAAGCTGCGACTGCCACACAGCTACATCGTCCTGTGCATTGAGAATAGCTTCATGACC 543

Qy   541 TCCTTCTCCAAGTAG 555
            |||||||| |||||
Db   544 TCTTTCTCCAAATAG 558

RESULT 7
US-09-775-325-3
; Sequence 3, Application US/09775325
; Patent No. 6500449
; GENERAL INFORMATION:
; APPLICANT: Abitbol, Marc
; APPLICANT: Uteza, Yves
; APPLICANT: Menasche, Maurice
; APPLICANT: Bossard, Carine
; APPLICANT: Van Den Berghe, Loic
; APPLICANT: Bonnel, Sebastian
; APPLICANT: Prats, Herve
; APPLICANT: Honiger, Jiri
; APPLICANT: Neuner-Jehle, Martin
; TITLE OF INVENTION: INTRAOCULAR TRANSPLANTATION OF ENCAPSULATED CELLS
; FILE REFERENCE: 8076.2020USD1
; CURRENT APPLICATION NUMBER: US/09/775,325
; CURRENT FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: US 09/449,293
; PRIOR FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Rattus rattus
US-09-775-325-3

Query Match   67.7%;   Score 375.8;   DB 4;   Length 558;
Best Local Similarity   79.8%;   Pred. No. 3e-66;
Matches 443;   Conservative   0;   Mismatches 112;   Indels   0;   Gaps   0;

Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCCCTGAACAGCCCGCAGCCG 60
            ||||| ||||||||||| |||||||||||||| |||||| ||||||| ||||||||||||
Db   4 CATACTCATCAGGACTTTCAGCCAGTGCTCCACCTGGTGGCACTGAACACCCCCCTGTCT 63

Qy   61 GGCGGCATGCGAGGCATCCGGGAGCGGACTTCCAGTGCTTCCAGCAGGCGCGCGCCGCG 120
            ||||||||| || |||||||||| ||||| ||||||| |||||||| ||| |||||||
Db   64 GGAGGCATGCGTGGTATCCGTGGCAGATTTCCAGTGCTTCCAGCAAGCCCGAGCCGTG 123

Qy   121 GGGCTGGCCCGCACCTTCCGGGCCTTCGTGTGTGCGGCTGCAGGACCTCTACAGCATC 180
            ||||| ||||||||||| ||||| |||||||| ||||| |||||| ||||||||||||
Db   124 GGGCTGTCGGGCACCTTCCGGGGCTTTCCTGTCCTCTAGGCTGCAGGATCTCTATAGCATC 183

Qy   181 GTGCGCGCGCCGACCGCACCGGGGTGCCGTCGTCGTCAACCTCAGGGACGAGGTGCTCTTC 240
            |||||| ||||||||| ||||||| ||||| |||||| ||||||||||||| ||||||||
Db   184 GTGCGCCGTGCTGCACCGGGGTCTGTGCCCATCGTCAACCTGAAGGACGAGGTGCTATCT 243

Qy   241 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGCCCGCATC 300
            ||||| |||| ||||| |||||| |||||| ||| ||||| |||||||||| ||||||
Db   244 CCCAGCTGGGACTCCCTGTTTTCTCGGCTCCCAGGGTCAACTGCAACCCGGGGCCCGCATC 303

Qy   301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCTGGCCCCGGAAGAGCGTGTGG 360
            ||||| |||||||| |||||||||||||| ||| |||||| |||||||||| ||||||
Db   304 TTTTCTTTCGACGGCAGAGATGTCCTGAGACACCACGACACCCACCCAGCGCCGTATGG 363

Qy   361 CACGGCTCCACCCCAGCCGGCGCCGCCTGACCAGCTACTGCGAGACGTGGCCGGACG 420
            ||||| ||||||||||||| | |||||| || |||| ||||| ||||||||||||||
Db   364 CACGGCTCGGACCCCAGTGGGCGGAGGCTGATGGAGCATGCTCGTGCTGTGGAGACT 423

Qy   421 GAGGCCCCGGCGGCCACCGGGGCAGGCGTCGTGCTGGCGGGCAGGCTGCTGGAGCAG 480
            |||||| |||| ||| ||||||| | ||||||||||| |||||||| ||||||||||
Db   424 GAAACTACTGGGGCTACAGTGCTCAGGCCTCTCCCTGCGTGTCCTGGAACAG 483

Qy   481 GAGGCCGCGAGCTCCCGCCACCGCCTTCGTGGTGCTCTGCATCGAGAACAGCGTCATGACC 540
            |||||||| ||||||||| |||||||| ||||||||||| ||||| |||||||||||||
Db   484 AAAGCTGCGACTGCCACACAGCTACATCGTCCTGTGCATTGAGAATAGCTTCATGACC 543

Qy   541 TCCTTCTCCAAGTAG 555
            |||||||| |||||
Db   544 TCTTTCTCCAAATAG 558

RESULT 8
US-09-775-174-3
; Sequence 3, Application US/09775174
; Patent No. 6663894
; GENERAL INFORMATION:
; APPLICANT: Abitbol, Marc
; APPLICANT: Uteza, Yves
; APPLICANT: Menasche, Maurice
; APPLICANT: Bossard, Carine
; APPLICANT: Van Den Berghe, Loic
; APPLICANT: Bonnel, Sebastian
; APPLICANT: Prats, Herve
; APPLICANT: Honiger, Jiri
; APPLICANT: Neuner-Jehle, Martin
; TITLE OF INVENTION: INTRAOCULAR TRANSPLANTATION OF ENCAPSULATED CELLS
; FILE REFERENCE: 8076.202USD2
; CURRENT APPLICATION NUMBER: US/09/775,174
; CURRENT FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: US 09/449,293
; PRIOR FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Rattus rattus
US-09-775-174-3

Query Match   67.7%; Score 375.8; DB 4; Length 558;
Best Local Similarity 79.8%; Pred. No. 3e-66;
Matches 443; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```
Qy    1 CACACCCACCAGGACTTCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG    60
          | ||| || ||||||| ||||| | || ||||||||||| | ||||||| |   |  |
Db    4 CATACTCATCAGGACTTTCAGCAGCTGCTCCACCTGGTGGCACTGAACACCCCCCGTGTCT 63

Qy   61 GGGGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGCGCGCGCCGCG  120
          || |||||||| | |||||| ||||| ||| |||||||||||||||| || || ||| |
Db   64 GGAGGCATGCGTGTATCCGTGGAGCAGATTTCCAGTGCTTCCAGCAAGCCCGAGCCGTG  123

Qy  121 GGGCTGGCGGCACCTTCCGGCGCCTTCCTGTCGTCGGGCTGCAGGACCTCACAGCATC  180
          ||||| | | ||||||||||  ||||||||||  | ||||||||||| ||| |||||||
Db  124 GGGCTGTCGGGCACCTTCCGGGCTTTCCTGTCCTCCGGGCTGCAGGATCTCTATAGCATC 183

Qy  181 GTGCGGCCGCGCCGACCGGCGGGGTCTGTGCCCATCGTCAACCTGAAGGACGAGGTGCTCTTC 240
          |||||||| ||| ||||| |||||||||||| ||||||||||||||| ||||||||||| |||
Db  184 GTGCGGCCGTGCTGACCGGGGGTCTGTGCCCATCGTCAACCTGAAGGACGAGGTGCTATCT   243

Qy  241 CCCAGTCTGGGAGGCCTTATTCTCGGGGTCCGAGGGCCAGCTGAAGGACGAGGCCCGCATC 300
          |||||| |||| ||||| |||| | |||| ||||||| || ||||  ||| ||||||||||
Db  244 CCCAGCTGGGACTCCCTGTTTTCTCGGCTCAACTGCAACCCCGGGGCCCCGAGCCCCGCATC 303

Qy  301 TTCTCTTCGACGGCAGAGATGTCCTGCAGCACCCCGGGCCCCGGAAGAGCGTGTGG  360
          |||| ||| |||| ||||||||||| |||| || | |||| ||| |||| | |||
Db  304 TTTTCTTTTGACGGCAGAGATGTCCTGAGACACCCAGGGCGCCAGGAGCGGTATGG   363

Qy  361 CACGGCTCCGACCCCAGCCCCGGCGCCTACTGCGACCACGCTGGCGGACG  420
          ||||| |||| ||||| |||| ||| |||| |||| ||||| |  |||
Db  364 CACGGCTCGGACCCCAGTGGGGCGGAGGCGTGCTGCTGGAGCAG   423

Qy  421 GAGGCCCGGCCGCCACGGGCCAGGCGTGCTCGGGGGCAGCAGCAG  480
          |||||||| || ||| || ||| || |||||| |||| ||| |||
Db  424 GAAACTACTGGGGCTCAGGTCAGGGGCGTCCCCGCTCCTGGAACAG 483

Qy  481 GAGGCCCGGCGAGCTGCCCGCCACAGCTGGTCATCGAGAACACGGTCATGACC 540
          |||||||| || |||| | |||||| || | | || |||| || | ||||||
Db  484 AAAGCTGCGAGCTGCCACAACAGCTACATCGTCCGTGTGCCATTGAGAATAGCTTCATGACC 543

Qy  541 TCCTTCCCAAGTAG  555
          || ||||| | |||
Db  544 TCTTTCCCAAATAG  558
```

RESULT 9
US-08-985-526-37
; Sequence 37, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 565 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-985-526-37

Query Match   67.7%; Score 375.8; DB 3; Length 565;
Best Local Similarity 79.8%; Pred. No. 3e-66;
Matches 443; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```
Qy    1 CACACCCACCAGGACTTCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG    60
          | ||| || ||||||| ||||| | || ||||||||||| | ||||||| |   |  |
Db   11 CATACTCATCAGGACTTTCAGCAGCTGCTCCACCTGGTGGCACTGAACACCCCCCGTGTCT 70

Qy   61 GGGGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGCGCGCGCCGCG  120
          || |||||||| | |||||| ||||| ||| |||||||||||||||| || || ||| |
Db   71 GGAGGCATGCGTGTATCCGTGGAGCAGATTTCCAGTGCTTCCAGCAAGCCCGAGCCGTG  130

Qy  121 GGGCTGGCGGCACCTTCCGGCGCCTTCCTGTCGTCGGGCTGCAGGACCTCACAGCATC  180
          ||||| | | ||||||||||  ||||||||||  | ||||||||||| ||| |||||||
Db  131 GGGCTGTCGGGCACCTTCCGGGCTTTCCTGTCCTCCGGGCTGCAGGATCTCTATAGCATC 190

Qy  181 GTGGCCCGCGCCGACCGGGGGTCGTCAACCTCAGGGATGTCCCGTCGTCAACCTGAAGGACGAGGTGCTCTTC 240
          ||||| || | ||||| |||||  || ||||| ||||| || ||  |||| |||||||||||||||||||
Db  191 GTGGCGCCGTGCTGACCGGGGGGTCGTCACGGGCGGGCCATCGTGCCATCGTCAACCTGAAGGACGAGGTGCTATCT 250

Qy  241 CCCAGTCTGGGAGGCCTTATTCTCGGGGTCCGAGGGCCAGCTGAAGGACGAGGCCCGCATC 300
          |||||| |||| ||||| |||| | |||| ||||||| || ||||  ||| ||||||||||
Db  251 CCCAGCTGGGACTCCCTGTTTTCTCGGCTCAACTGCAACCCCGGGGCCCCGAGCCCCGCATC 310

Qy  301 TTCTCTTCGACGGCAGAGATGTCCTGCAGCACCCCGGGCCCCGGAAGAGCGTGTGG  360
          |||| ||| |||| ||||||||||| |||| || | |||| ||| |||| | |||
Db  311 TTTTCTTTTGACGGCAGAGATGTCCTGAGACACCCAGGGCGCCAGGAGCGGTATGG   370

Qy  361 CACGGCTCCGACCCCAGCCCCGGCGCCTACTGCGACCACGCTGGCGGACG  420
          ||||| |||| ||||| |||| ||| ||  |||| ||||| |  |||
Db  371 CACGGCTCGGACCCCAGTGGGCGGAGGCGTGATGGAGAGTTACTGTGAGACATGGGCAACT 430
```

Qy 421 GAGGCCCGGGCGGCCACCGGGCCAGGCGGCGTCGTCGCTCGGCGGGCAGGCTGCTGGAGCAG 480
         |||||| |||||||||| ||| | | | | ||| ||||| |||||| |||||||
Db 431 GAAACTACTGGGGCTACAGGTCAGGTCAGGCTCCTCCCGCTGTCTCAGGGAGGCTTCCTGGAACAG 490

Qy 481 GAGGCCCGAGTGCGCCACGCCCTTCGTCGTGGTGCTCTGCATCGAGAACAGCGTCATGACC 540
         ||||| || ||||| || |||||| | ||||| |||| |||||||| ||| |||||||
Db 491 AAAGCTGCGAGCTGCCACAACAGCTACATCGTCCTGTGCATTGAGAATAGCTTCATGACC 550

Qy 541 TCCTTCTCCAAGTAG 555
         || ||||||||| ||
Db 551 TCTTTCTCCAAATAG 565

RESULT 10
US-08-159-784-1
; Sequence 1, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
;   APPLICANT: Bjorn R. Olsen
;   TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
;   NUMBER OF SEQUENCES: 9
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: Massachusetts
;   COUNTRY: U.S.A.
;   ZIP: 02110-2804
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;   COMPUTER: IBM PS/2 Model 50Z or 55SX
;   OPERATING SYSTEM: MS-DOS (Version 5.0)
;   SOFTWARE: WordPerfect (Version 5.1)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/159,784
;   FILING DATE: December 1, 1993
;   CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: John F. Freeman
;   REGISTRATION NUMBER: 29,066
;   REFERENCE/DOCKET NUMBER: 00246/170001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 542-5070
;   TELEFAX: (617) 542-8906
;   TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 4031
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-08-159-784-1

Query Match   67.4%; Score 374.2; DB 1; Length 4031;
Best Local Similarity   79.6%; Pred. No. 7e-66;
Matches 442; Conservative   0; Mismatches 113; Indels   0; Gaps   0;

Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCTGCACCTGGTGCTGCAAGAGCAGCCCGAACAGCCCGAGCCG   60
          |||| |||| |||||||| |||||||| |||||| ||||||||| |||||| ||| || ||| ||||||||||||
Db 3313 CATACTCATCATCAGGACTTTCAGCCAGTGCTCCACCTGGTGGCACTGGTGGCACTGAACACCCCCCCTGTCT 3372

Qy   61 GGGCGGCATGCGAGGCATCCGGGAGGCGGACTTCCAGTGCTTCCAGCAGGCGCGGCCGCG 120
          |||||||||||| || |||||| |||| || |||||||||||||||||| | |||||||
Db 3373 GGAGGCATGCGTGGTATCCGTGGAGCAGATTTCCAGTGCTTCCAGCAGCCCGAGCCGTG 3432

Qy   121 GGGCTGGCCGGCACCTTCCGGGCCTTCCTGTGTCGCGCGGCTGCAGGACCTCACAGCATC 180
          |||||| | |||||||||| ||| ||||| | | ||| |||||||| | |||||||
Db 3433 GGGCTGTGCGGGCACCTTCCGGGCGTTTCCTCTAGGCTGCCAGGATCTCATAGCATC 3492

Qy   181 GTGCGCGCGCGACCGCACCGGGGTGCCCGTCGCAACCTCAGGACGAGGTGCTCTTC 240
          ||||||| ||| ||||| |||||| |||| |||| ||| ||||| | ||||
Db 3493 GTGCGCCGTGCTGCTGACCGCGGGGTCTGTGCCCATCGTGCCCATCGTCAACCTGACGAAGGACGAGAGGTGCTATCT 3552

Qy   241 CCCAGCTGGGAGGCCTTATTCTCCGGGCTCCGAGGGCCAGCTGAGTACCCGGGCCGAGCCCGGGCCATC 300
          |||||||||| || | | |||| ||| |||||| |||| ||||||| |||||||||||||||
Db 3553 CCCAGCTGGGACTCCCGTGTTTCTCGGCTCCCAGGGTACCTCAGGTCAAGTGCAACCCGGGCCGCATC 3612

Qy   301 TTCTCTTCGACGGCAGAGATGTCTCGCAGCACCCGCCCGGCCCCGGAAGAGCGTGTGG 360
          ||| ||| ||||||||||||||| | || ||||| || || || ||||| ||||
Db 3613 TTTTCTTTGACGGCAGAGATGTCCTGAGACACCCAGCCTGGCCGCAGAAGAGGTATGG 3672

Qy   361 CACGGGCTCCGACCCCAGCGCGCGCGGGCGCCGCCTGACCAGCTACTCCGAGACGTGGCGGACG 420
          ||||||||| ||||||| ||| || | || ||||||| |||| ||||| |||| |||
Db 3673 CACGGGCTCGGACCCCAGTGGGCGGAGGCTCGATGGAGGCGTTACTGTGAGACATGGCGAACT 3732

Qy   421 GAGGCCCCGGGCGGCCACCGGCCAGCGGGCGTCGTCGCTCGGCGGGCAGGCTGCTGGAGCAG 480
          || | |||||||| ||| | | | | ||| ||||| |||||| |||||||
Db 3733 GAAACTACTGGGGCTACAGGTCAGGTCAGGCTCCTCCCGCTGTCTCAGGGAGGCTTCCTGGAACAG 3792

Qy   481 GAGGCCCGAGTGCGCCACGCCCTTCGTCGTGGTGCTCTGCATCGAGAACAGCGTCATGACC 540
          ||||| || ||||| || |||||| | ||||| |||| |||||||| ||| |||||||
Db 3793 AAAGCTGCGAGCTGCCACAACAGCTACATCGTCCTGTGCATTGAGAATAGCTTCATGACC 3852

Qy   541 TCCTTCTCCAAGTAG 555
          || ||||||||| ||
Db 3853 TCTTTCTCCAAATAG 3867

RESULT 11
US-09-561-500-12
; Sequence 12, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
;   APPLICANT: Philip E. Thorpe
;   APPLICANT: Rolf A. Brekken
;   TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
;   FILE REFERENCE: 4001.002500
;   CURRENT APPLICATION NUMBER: US/09/561,500
;   CURRENT FILING DATE: 2000-04-28
;   PRIOR APPLICATION NUMBER: 60/131,432
;   PRIOR FILING DATE: 1999-04-28
;   NUMBER OF SEQ ID NOS: 44
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 573
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
;   OTHER INFORMATION: OLIGONUCLEOTIDE
;   NAME/KEY: CDS
;   LOCATION: (1)..(573)
US-09-561-500-12

Query Match   67.3%; Score 373.4; DB 4; Length 573;
Best Local Similarity   79.9%; Pred. No. 8.9e-66;
Matches 440; Conservative   0; Mismatches 111; Indels   0; Gaps   0;

Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCTGCACCTGGTGCTGCAAGAGCAGCCCGAACAGCCCGAGCCG   60
         |||| |||| |||||||| |||||||| |||||| ||||||||| |||||| ||| || ||| ||||||||||||
Db   22 CATACTCATCATCAGGACTTTCAGCCAGTGCTCCACCTGGTGGCACTGGTGGCACTGAACACCCCCCCTGTCT   81

Qy   61 GGGCGGCATGCGAGGCATCCGGGAGGCGGACTTCCAGTGCTTCCAGCAGGCGCGGCCGCG 120
         |||||||||||| || |||||| |||| || |||||||||||||||||| | |||||||
Db   82 GGAGGCATGCGTGGTATCCGTGGAGCAGATTTCCAGTGCTTCCAGCAGCCCGAGCCGTG 141

Qy 121 GGGCTGGCCGGCACCTTCCGTGCGCGGCTTGCGCGGCCTTGCAGGACCTGCAGGACATC 180
         |||||| | |||||||||| ||| || | | ||| |||||||| | |||||||
Db 142 GGGCTGTGCGGGCACCTTCCGGGCGTTTCCTCTAGGCTGCCAGGATCTCATAGCATC 201

Qy 181 GTGCGGCCCGCGCGACCGCACCGGGGTGCCCGTGCCGTGCTCTTC 240
         ||||| || || ||| |||||||||| |||| |||||| ||||

RESULT 12
US-09-561-108-12
; Sequence 12, Application US/09561108
; Patent No. 6342221
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/561,108
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: OLIGONUCLEOTIDE
; NAME/KEY: CDS
; LOCATION: (1)..(573)
US-09-561-108-12

Query Match   67.3%;   Score 373.4;   DB 4;   Length 573;
Best Local Similarity 79.9%;   Pred. No. 8.9e-66;
Matches 440; Conservative   0; Mismatches 111; Indels 0; Gaps 0;

Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCGTGCACCTGGCCCTGGAACAGCCCGCAGCCG   60
Db   22 CATACTCATCAGGACTTTCAGCCAGTGCTTCCACCTGGTGGGCACTGGAACACCCCCCTGTCT   81
Qy   61 GGCGGCATGCGGAGGCATCCGGGAGCGGACTTCCAGTGCTTCCAGCGAGCGCGCGCGCG   120
Db   82 GGAGGCATGCGTGATCCGTGAGCAGATTTCCAGTGCTTCCAGCAAGCCCGAGCCGTG   141
Qy   121 GGGCTGGCCGCGCACCTTCCGGGCCTTCGTCGTGCGGCTGCAGCACCTTCACAGCATC   180
Db   142 GGGCTGTCGGGCACCTTCCGGGCTTTCCTGTCCTCTAGGCTGCAGGATCTTCTATAGCATC   201
Qy   181 GTGCGCCGCGCCGACCGGGCGACCGGGGGTCTGTCGTGCCCATCGTCGTGCCCGTGCTATCT   240
Db   202 GTGCGCCGTGCTGACCGGGGGTCTGTGCCCATCGTCGTGCCCATCGTCAAGGAGGAGGTGCTATCT   261
Qy   241 CCCAGCTGGGAGGCCTTATTCTCGGCGTCCGAGGGGCAGCCCGGGGCCCGCATC   300

Db   202 GTGCGCCGTGCTGCTGACCGGGGGTCTGTGCCCATCGTCGTGCCCATCGTCAAGGAGGAGGTGCTATCT   261
Qy   241 CCCAGCTGGGAGGCCTTATTCTCGGCGTCCGAGGGGCAGCCCGGGGCCCGCATC   300
Db   262 CCCAGCTGGGACTCCCTGTTTTCTGGCTCCCAGGGTCAACTGCAACCCGGGGCCCGCATC   321
Qy   301 TTCTCCTTCGACGGCAGAGATGTCCTCGCAGCACCCCGCCTGGCCCCCGAGAGAGCGTGTGG   360
Db   322 TTTTCTTTTGACGGCAGAGATGTCCTGAGACACCCAGCCTGGCCCGCAGAGAGCGTATGG   381
Qy   361 CACGGCTCCGACCCCAGCGGCCGCCTGCGACAGCTACTGCCGAGACGTGGCGGACG   420
Db   382 CACGGCTCGGACCCCAGTGGCGCCGGAGGCTGATGGAGCTGATGCGTGCGGCAGGTTACTGCGG   441
Qy   421 GAGGCCCGGCGCCACCGGCGGTCGTGCTGCTGGTGCGGCGGCAGGCTGCTGGAGCAG   480
Db   442 GAAACTACTGGGGCTACAGGTCAGGCCTCAGGCCTCCTCGTCGTGCTGCTCCTGGAACAG   501
Qy   481 GAGGCCCGAGCTGCGCCCCAGCCCTTCGTGTGCTCCTGTGCTGCATCGAGAACAGCGTCATGACC   540
Db   502 AAAGCTGCGAGCTGCCACAACAGCTACATCGTCCTGTGCATTGAGAATAGCTTCATGACC   561
Qy   541 TCCTTCTCCAA   551
Db   562 TCTTTCTCCAA   572

RESULT 13
US-09-561-526-12
; Sequence 12, Application US/09561526
; Patent No. 6416758
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002586
; CURRENT APPLICATION NUMBER: US/09/561,526
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: OLIGONUCLEOTIDE
; NAME/KEY: CDS
; LOCATION: (1)..(573)
US-09-561-526-12

Query Match   67.3%;   Score 373.4;   DB 4;   Length 573;
Best Local Similarity 79.9%;   Pred. No. 8.9e-66;
Matches 440; Conservative   0; Mismatches 111; Indels 0; Gaps 0;

Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCGTGCACCTGGCCCTGAACAGCCCGCAGCCG   60
Db   22 CATACTCATCAGGACTTTCAGCCAGTGCTTCCACCTGGTGGGCACTGAACACCCCCCTGTCT   81
Qy   61 GGCGGCATGCGGAGGCATCCGGGAGCGGACTTCCAGTGCTTCCAGCGAGCGCGCGCG   120
Db   82 GGAGGCATGCGTGATCCGTGAGCAGATTTCCAGTGCTTCCAGCAAGCCCGAGCCGTG   141
Qy   121 GGGCTGGCCGCGCACCTTCCGGGCCTTCGTCGTGCGGCTGCAGCACCTTCACAGCATC   180
Db   142 GGGCTGTCGGGCACCTTCCGGGCTTTCCTGTCCTCTAGGCTGCAGGATCTCTATAGCATC   201
Qy   181 GTGCGCCGCGCCGACCGGGCGACCGGGGGTCTGTCGTGCCCGTGCTTCGTCGGACGAGAGTGCTCTTC   240
Db   202 GTGCGCCGTGCTGCTGACCGGCCGGGGGTCTGTGCATCGTCGTGCCCATCGTCAAGGAGGAGGTGCTATCT   261
Qy   241 CCCAGCTGGGAGGCCTTATTCTCGGCGTCCGAGGGGCAGCCCGGGGCCCGCATC   300

Qy   241 CCCAGCTGGGAGGCCTTATTCTCGGCGTCCGAGGGGCAGCCCGGGGCCCGCATC   300
Db   262 CCCAGCTGGGACTCCCTGTTTTCTGGCTCCCAGGGTCAACTGCAACCCGGGGCCCGCATC   321
Qy   301 TTCTCCTTCGACGGCAGAGATGTCCTCGCAGCACCCCGCCTGGCCCCCGAGAGAGCGTGTGG   360
Db   322 TTTTCTTTTGACGGCAGAGATGTCCTGAGACACCCAGCCTGGCCCGCAGAGAGCGTATGG   381
Qy   361 CACGGCTCCGACCCCAGCGGCCGCCTGCGACAGCTACTGCCGAGACGTGGCGGACG   420
Db   382 CACGGCTCGGACCCCAGTGGCGCGGAGGCTGATGGAGCTGATGCGTGCGGCAGGTTACTGCGGAACT   441
Qy   421 GAGGCCCGGCGCCACCGGCGGTCGTGCTGCTGGTGCGGCGGCAGGCTGCTGGAGCAG   480
Db   442 GAAACTACTGGGGCTACAGGTCAGGCCTCAGGCCTCCTCGTCGTGCTGCTCCTGGAACAG   501
Qy   481 GAGGCCCGAGCTGCGCCCCAGCCCTTCGTGTGCTCCTGTGCTGCATCGAGAACAGCGTCATGACC   540
Db   502 AAAGCTGCGAGCTGCCACAACAGCTACATCGTCCTGTGCATTGAGAATAGCTTCATGACC   561
Qy   541 TCCTTCTCCAA   551
Db   562 TCTTTCTCCAA   572

; Sequence 12, Application US/09561499
; Patent No. 6524583
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/09/561,499
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: OLIGONUCLEOTIDE
; NAME/KEY: CDS
; LOCATION: (1)..(573)
US-09-561-499-12

Query Match   67.3%; Score 373.4; DB 4; Length 573;
Best Local Similarity 79.9%; Pred. No. 8.9e-66;
Matches 440; Conservative   0; Mismatches 111; Indels   0; Gaps   0;

```
Qy    1 CACACCCACCAGGACTTCCAGCTGGTCGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG   60
        ||||| ||| |||||||| ||||  ||| ||||| ||||||| | ||| |||||  |||   |
Db   22 CATACTCATCAGGACTTTCAGCCAGTGCTCCACCTGGTGGCACTGAACACCCCCCTGTCT    81

Qy   61 GGCGGCATGCGAGGCATCCGGGAGGCGGACTTCCAGTGCTTCCAGCAGCGGCGGCGGCCGCG   120
        || ||||||||  || ||| || |||||| | ||||||||||||||||  | || |||||||
Db   82 GGAGGCATGCGGTGGTATCCCGTGGAGCAGATTTCAGTGCTTCCAGCAGCCCGAGCCGTG   141

Qy  121 GGGCTGGCCGGCGACCTTCCGGGCCTTCGTGTCGTGCGGCTGCAGGACCTTCTACAGCATC   180
        |||||| | |||| ||||||||||| || || |  || ||||||||| | || |||||||||
Db  142 GGGCTGTCGGGCACCTTCCGGGCTTTCCTGTCTCTCTAGGCTGCAGGATCTCTATAGCATC   201

Qy  181 GTGCGCCGGCGCCGACCAACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC   240
        |||||||| | || | || ||||||| | |||| |||||||| |||||||||||| | |
Db  202 GTGCGCCGGTGCTGACCGACCGGGGGTCTGTGCCCATCGTCAACCTGAAGGACGAGGTGCTATCT   261

Qy  241 CCCAGCTGGGAGGGCCTTATTCTCCGGGGCTGCGAGGGCCAGCTGAAAGCCCGGGCCCGCATC   300
        |||||||||||| | || ||| |||  |  || ||||||||||||||||||||  ||||||||
Db  262 CCCAGCTGGGAGCTCCCTGTTTCTCGGCTCCACCGGGTCAACTGCAACCGGGGCCCCGCATC   321
```

; Sequence 12, Application US/09998831
; Patent No. 6676941
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
; TITLE OF INVENTION: INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/998,831
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 09/561,108
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: OLIGONUCLEOTIDE
; NAME/KEY: CDS
; LOCATION: (1)..(573)
US-09-998-831-12

Query Match   67.3%; Score 373.4; DB 4; Length 573;
Best Local Similarity 79.9%; Pred. No. 8.9e-66;
Matches 440; Conservative   0; Mismatches 111; Indels   0; Gaps   0;

```
Qy    1 CACACCCACCAGGACTTCCAGCTGGTCGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG   60
        ||||| ||| |||||||| ||||  ||| ||||| ||||||| | ||| |||||  |||   |
Db   22 CATACTCATCAGGACTTTCAGCCAGTGCTCCACCTGGTGGCACTGAACACCCCCCTGTCT    81

Qy   61 GGCGGCATGCGAGGCATCCGGGAGGCGGACTTCCAGTGCTTCCAGCAGCGGCGGCGGCCGCG   120
        || ||||||||  || ||| || |||||| | ||||||||||||||||  | || |||||||
Db   82 GGAGGCATGCGGTGGTATCCCGTGGAGCAGATTTCAGTGCTTCCAGCAGCCCGAGCCGTG   141

Qy  121 GGGCTGGCCGGCGACCTTCCGGGCCTTCGTGTCGTGCGGCTGCAGGACCTTCTACAGCATC   180
        |||||| | |||| ||||||||||| || || |  || ||||||||| | || |||||||||
Db  142 GGGCTGTCGGGCACCTTCCGGGCTTTCCTGTCTCTCTAGGCTGCAGGATCTCTATAGCATC   201

Qy  181 GTGCGCCGGCGCCGACCAACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC   240
        |||||||| | || | || ||||||| | |||| |||||||| |||||||||||| | |
Db  202 GTGCGCCGGTGCTGACCGACCGGGGGTCTGTGCCCATCGTCAACCTGAAGGACGAGGTGCTATCT   261

Qy  241 CCCAGCTGGGAGGGCCTTATTCTCCGGGGCTGCGAGGGCCAGCTGAAAGCCCGGGCCCGCATC   300
        |||||||||||| | || ||| |||  |  || ||||||||||||||||||||  ||||||||
Db  262 CCCAGCTGGGAGCTCCCTGTTTCTCGGCTCCACCGGGTCAACTGCAACCGGGGCCCCGCATC   321

Qy  301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCGACCCTGGCCCGGAGAGAGGCGTGTGG   360
```

```
Db  301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCGACCCTGGCCCGGAGAGAGGCGTGTGG   360
Qy  322 TTTTCTTTTGACGGCAGAGATGTCCTGCAGACACCCAGCCTGGCCGCAGAGAGCGTATGG   381
Db  361 CACGGCTCCGACCCCAGCCGGCGCCGCCTGACCGACAGCTACTGCGAGACGTGGCGGACG   420
Qy  382 CACGGCTCGGACCCCAGTGGCGGAGGCTGATGGAGACGTTACTGTGAGACATGGCGAACT   441
Db  421 GAGGCCCCGGCGGCCACCGGCAGGCGTCGTCGCTGCGCTCCTCCCTGCGTGTCGGAGCAG   480
Qy  442 GAAACTACTGGGCTACAGGTCAGGCCTTCGTGTCTGTCGCTGGGGCAGGCTCCTGGAACAG   501
Db  481 GAGGCCGCGAGCTGCCGCCACGCCTTCGGTGCTCTGCATCGAGAACAGCGTCATGACC   540
Qy  502 AAAGTGCGAGCTGCCACAACAGCTACATCGTCCTGTGCATTGAGAATAGCTTCATGACC   561
Db  541 TCCTTCTCCAA   551
Qy  562 TCTTTCTCCAA   572
```

```
Db     322  |||||||| ||||||||||||||||||| |||||||||||||| ||||||||| | ||||||||||| |||
            TTTTCTTTTGACGGCAGAGAGATGTCCTGAGACACCCAGCCTGGCCGCAGAGAGCGTATGG  381

Qy     361  CACGGCTCCGACCCCAGCGCGGGCGCGCCTGACCGACAGCTACTGCGAGACGTGGCGGACG  420
            ||||||||| |||||| ||||| | |||||| ||| ||| |||| |||| ||| |
Db     382  CACGGCTCGGACCCCAGTGGGCGGAGGCTGATGGAGAGTTACTGTGAGACATGGCGAACT  441

Qy     421  GAGGCCCCGGCGGCCACCGGGCAGGCGTCGTCGCTGCTGGCGGGGCAGGCTGCTGGAGCAG  480
            |||||||| | | ||||||| ||||| |||||||| |||||| ||
Db     442  GAAACTACTGGGGCTACAGGTCAGGCCTCCCGCCTGTCGCAGGCAGGCTCCTGGAACAG  501

Qy     481  GAGGCCGGGAGCTGCGCGCCACGCCTTCGTGGTGCTCTGCATCGAGAACAGCGTCATGACC  540
            |||||||||| ||||| ||| |||| |||| ||||| ||| ||||||
Db     502  AAAGCTGCGAGCTGCCACACAGCTACATCGTCCTGTGCATTGAGAATAGCTTCATGACC  561

Qy     541  TCCTTCTCCAA  551
            || |||||||||
Db     562  TCTTTCTCCAA  572
```

Search completed: March 29, 2004, 13:41:22
Job time : 59.1416 secs

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        March 29, 2004, 13:38:55 ; Search time 218.15 Seconds
                                        (without alignments)
                                        9471.775 Million cell updates/sec

Title:      US-09-938-391-3
Perfect score: 555
Sequence:   1 cacaccaccaggactcca.........tgacctcttcctccaagtag 555

Scoring table:  IDENTITY_NUC
           Gapop 10.0 , Gapext 1.0

Searched:   245896 seqs, 181504845 residues

Total number of hits satisfying chosen parameters:    4917892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Published_Applications_NA:*
    1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
    2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
    3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
    4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
    5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
    6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
    7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
    8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
    9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
   10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
   11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
   12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
   13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
   14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
   15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
   16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
   17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
   18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	555	100.0	555	10	US-09-938-391-3	Sequence 3, Appli
2	555	100.0	829	10	US-09-938-391-1	Sequence 1, Appli
3	550.4	99.2	552	14	US-10-131-241-50	Sequence 50, Appl
4	550.4	99.2	552	14	US-10-292-418-34	Sequence 34, Appl
5	418	75.3	546	14	US-10-042-347-4	Sequence 4, Appli
6	418	75.3	549	14	US-10-131-241-53	Sequence 53, Appli
7	418	75.3	549	14	US-10-292-418-3	Sequence 3, Appl
8	418	75.3	552	9	US-09-873-676-30	Sequence 30, Appl
9	418	75.3	3394	14	US-09-880-107-2178	Sequence 2178, Ap
10	418	75.3	4551	14	US-10-060-036-144	Sequence 144, App
11	418	75.3	4875	15	US-10-264-049-835	Sequence 835, App
12	417	75.1	551	13	US-10-080-797-2	Sequence 2, Appli
13	415.6	74.9	632	14	US-10-131-241-51	Sequence 51, Appl
14	411.6	74.2	555	12	US-10-210-172-161	Sequence 161, App
15	410.8	74.0	534	14	US-10-042-347-6	Sequence 6, Appli

```
16 410.8 74.0 537 14 US-10-131-241-59    Sequence 59, Appl
17 403   72.6 540 14 US-10-131-241-48    Sequence 48, Appl
18 375.8 67.7 558 9  US-09-775-174-3     Sequence 3, Appli
19 375.8 67.7 558 9  US-09-775-325-3     Sequence 3, Appli
20 375.8 67.7 565 13 US-10-036-869-37    Sequence 37, Appl
21 374.2 67.4 624 13 US-10-080-797-4     Sequence 4, Appli
22 374.2 67.4 624 14 US-10-422-934-70    Sequence 70, Appl
23 373.4 67.3 573 9  US-09-998-831-12    Sequence 12, Appl
24 373.4 67.3 573 14 US-10-373-561-12    Sequence 12, Appl
25 371.8 67.0 552 14 US-10-292-418-17    Sequence 17, Appl
26 315.2 56.8 482 12 US-10-210-172-163   Sequence 163, App
27 302.8 54.6 480 12 US-10-210-172-165   Sequence 165, App
28 227.2 40.9 306 15 US-10-430-503-37    Sequence 37, Appl
29 187.8 33.8 574 14 US-10-060-036-64    Sequence 64, Appl
30 171.2 30.8 5204 10 US-09-371-392-159  Sequence 159, App
31 171.2 30.8 5222 15 US-10-113-937-67   Sequence 67, Appl
32 171.2 30.8 5222 15 US-10-159-563-67   Sequence 67, Appl
33 95.4  17.2 900 14 US-10-131-241-45    Sequence 45, Appl
34 74.4  13.4 434 14 US-10-060-036-4     Sequence 4, Appli
35 58.8  10.6 2616 15 US-10-369-493-42226 Sequence 42226, A
36 58.4  10.5 1728 14 US-10-156-761-4493  Sequence 4493, Ap
37 58.4  10.5 9025608 14 US-10-156-761-1   Sequence 1, Appli
38 57    10.3 1548 14 US-10-156-761-3078  Sequence 3078, Ap
39 57    10.3 9025608 14 US-10-156-761-3078 Sequence 1, Appli
40 56.2  10.1 2079 14 US-10-076-816-3     Sequence 3, Appli
41 56.2  10.1 2079 14 US-10-077-381-3     Sequence 3, Appli
42 56.56 10.1 1785 14 US-10-156-761-7479  Sequence 7479, Ap
43 55.6  10.1 1821 14 US-10-156-761-3274  Sequence 3274, Ap
44 55.6  10.0 1704 14 US-10-156-761-3928  Sequence 3928, Ap
45 55.2  9.9  1392 14 US-10-214-446-7     Sequence 7, Appli
```

ALIGNMENTS

```
RESULT 1
US-09-938-391-3
; Sequence 3, Application US/09938391
; Publication No. US2003015809A1
; GENERAL INFORMATION:
;  APPLICANT: Tong, et al.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND TREATING
;  TITLE OF INVENTION:  DISORDERS INVOLVING ANGIOGENESIS
;  FILE REFERENCE: PC10790A
;  CURRENT APPLICATION NUMBER: US/09/938,391
;  CURRENT FILING DATE:  2001-08-24
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;  LENGTH: 555
;  TYPE: DNA
;  ORGANISM: CANINE ENDOSTATIN NUCLEOTIDE SEQUENCE
US-09-938-391-3
```

```
Query Match           100.0%;  Score 555;  DB 10;  Length 555;
Best Local Similarity 100.0%;  Pred. No. 1.2e-117;
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy    1 CACACCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 CACACCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG   60

Qy   61 GGGGCATGCGAGGCATCCGGGAGGCGGACTTCCAGTGCTTCCAGTGCTTCCAGCAGGCGCGGCGCGCCGCG  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGGGCATGCGAGGCATCCGGGAGGCGGACTTCCAGTGCTTCCAGTGCTTCCAGCAGGCGCGGCGCGCCGCG  120

Qy  121 GGGCTGGCCGGCACCTTCCGGGCCTTCCGGGCCTTCCGTCGTCGCGGCTGCAGGACCTCTACAGCATC  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GGGCTGGCCGGCACCTTCCGGGCCTTCCGGGCCTTCCGTCGTCGCGGCTGCAGGACCTCTACAGCATC  180

Qy  181 GTGCGCCGCGCCGACGGCACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGAGGTGCTCTTC  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GTGCGCCGCGCCGACGGCACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGAGGTGCTCTTC  240
```

Left column

```
Qy  241  CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC  300
         ||||||||||||||||| |||||||||||||||||||||||||||||||||||| |||||
Db  241  CCCAGCTGGGAGGCCTTTATTCTGGGGCTTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC  300

Qy  301  TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCTCCTGGGCCCCGGAAGAGCGTGTGG  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCTCCTGGGCCCCGGAAGAGCGTGTGG  360

Qy  361  CACGGCTCCGACCCCAGCCGGGCCGGCCGCCTGACCAGCTACTGCGGGCACGTGGCGGACG  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  CACGGCTCCGACCCCAGCCGGGCCGGCCGCCTGACCAGCTACTGCGAGACGTGGCGGACG  420

Qy  421  GAGGCCCCGGCCGGCCACCGGCAGGGTCGTGCTGCTGGGCGGCAGCGTGCTGGAGCAG  480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  GAGGCCCCGGCCGGCCACCGGCAGGGTCGTGCTGCTGGGCGGCAGCGTGCTGGAGCAG  480

Qy  481  GAGGCCGCGAGTCGCCGCCACGCCTTCGTGGTGCTCGTGCTGCTCATGGACG  540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  GAGGCCGCGAGTCGCCGCCACGCCTTCGTGGTGCTCGTGCTGCTCATGGACG  540

Qy  541  TCCTTCTCCAAGTAG  555
         |||||||||||||||
Db  541  TCCTTCTCCAAGTAG  555
```

RESULT 2
US-09-938-391-1
; Sequence 1, Application US/09938391
; Publication No. US20030158099A1
; GENERAL INFORMATION:
; APPLICANT: Tong, et al.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND TREATING
; TITLE OF INVENTION: DISORDERS INVOLVING ANGIOGENESIS
; FILE REFERENCE: PC10790A
; CURRENT APPLICATION NUMBER: US/09/938,391
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 829
; TYPE: DNA
; ORGANISM: CANINE PRO ENDOSTATIN NUCLEOTIDE SEQUENCE
US-09-938-391-1

```
Query Match          100.0%;  Score 555;  DB 10;  Length 829;
Best Local Similarity 100.0%;  Pred. No. 1.2e-117;
Matches 555;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    1  CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCCCTGAACAGCCCGCAGCCG   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  139  CACACCCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCCCTGAACAGCCCCAGCCG  198

Qy   61  GGCGGCATGCGAGGCATCCGGGAGGCCGACTTCCAGTGCTTCCAGCAGGCCGCGGCCGCG  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  GGCGGCATGCGAGGCATCCGGGAGGCCGACTTCCAGTGCTTCCAGCAGGCCGCGGCCGCG  120

Qy  121  GGGCTGGCGCGGCACCTTCCGGGCCTTCCGTGTCGTCGGCCTGCGTCAGGACGAGGTGCTTTC  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  199  GGGCTGGCGCGGCACCTTCCGGGCCTTCCGTGTCGTCGGCCTGCGTCAGGACGAGGTGCTTTC  258

Qy  121  GGGCTGGCCGGCAGCCTTCCGGGCCTTCCGGGCGTGCCGGTGCAACCTCAGGGACGAGGTGCTCTTC  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  259  GGGCTGGCCGGCAGCCTTCCGGGCCTTCCGGGCGTGCCGGTGCAACCTCAGGGACGAGGTGCTCTTC  318

Qy  181  GTGCGCGCGGCCGCACCGGCCGCACCGGGTGCCCGGTGCCCGGTCAACCTCAGGGGACGAGGTGCTCTTC  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  GTGCGCGCGGCCGCACCGGCCGCACCGGGTGCCCGGTGCCCGGTCAACCTCAGGGGACGAGGTGCTCTTC  240

Qy  241  CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  319  CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC  378

Qy  301  TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCTCCTGGGCCCCGGAAGAGCGTGTGG  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  379  CCCAGCTGGGAGGCCTTATTCTCGGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC  438

Qy  301  TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCTCCTGGGCCCCGGAAGAGCGTGTGG  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCTCCTGGGCCCCGGAAGAGCGTGTGG  360

Qy  439  TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCTCCTGGGCCCCGGAAGAGCGTGTCG  498
```

Right column

```
Qy  361  CACGGCTCCGACCCCAGCGGGCGGGCCGCCTGACCGACAGCTACTGCGAGCACGTGGCGGACG  420
         |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db  499  CACGGCTCCGACCCCAGCGGGCGCGCCGCCTGACCGACAGCTACTGCGAGACGTGGCGGACG  558

Qy  421  GAGGCCCCGGCGGCCACCGGCAGGGTCGTGCTGCTGGGCGGCAGCGTGCTGGAGCAG  480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  559  GAGGCCCCGGGCGGCCACCGGCAGGGTCGTGCTGCTGGGCGGCAGCGTGCTGGGACAG  618

Qy  481  GAGGCCGCGAGTCGCCGCCACGCCTTCGTGGTGCTCGTGCTGCTCATGGAGAACAGCAGGGTCATGACC  540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  619  GAGGCCGCGAGTCGCCGCCACGCCTTCGTGGTGCTCGTGCTGCTCATCGAGAACAGCGTCATGACC  678

Qy  541  TCCTTCTCCAAGTAG  555
         |||||||||||||||
Db  679  TCCTTCTCCAAGTAG  693
```

RESULT 3
US-10-131-241-50
; Sequence 50, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolife
; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
; FILE REFERENCE: 05213-0344 43170-271555
; CURRENT APPLICATION NUMBER: US/10/131,241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; SOFTWARE: PatentIn version 3.1
; NUMBER OF SEQ ID NOS: 65
; SEQ ID NO 50
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Canine sp.
US-10-131-241-50

```
Query Match          99.2%;   Score 550.4;  DB 14;  Length 552;
Best Local Similarity 99.8%;   Pred. No. 1.4e-116;
Matches 551;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

Qy    1  CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCCCTGAACAGCCCGCAGCCG   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCCCTGAACAGCCCGCAGCCG   60

Qy   61  GGCGGCATGCGAGGCATCCGGGAGGCCGACTTCCAGTGCTTCCAGCAGGCGCGGCCGCG  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  GGCGGCATGCGAGGCATCCGGGAGGCCGACTTCCAGTGCTTCCAGCAGGCGCGGCCGCG  120

Qy  121  GGGCTGGCGCGGCACCTTCCTGGGCCTGCGTGTCGTCGGCCTGCGTCAGGACCTCTACAGCATC  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  GGGCTGGCCGGCACCTTCCTGGGCCTGCGTGTCGTCGGCCTGCGTCAGGACCTCTACAGCATC  180

Qy  181  GTGCGCCGCGGCCGCACCGGGCGGGTGCCGTCAACCTCAGGGACGAGGTGCTCTTC  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  GTGCGCCGCGGCCGCACCGGGCGGGTGCCGTCAACCTCAGGGACGAGGTGCTCTTC  240

Qy  241  CCCAGCTGGGAGGCCTATTCTCGGGGTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  CCCAGCTGGGAGGCCTTATTCTCGGGGTCCGAGGGCCAGCCAGCCAGCCCGGGGCCCCGCATC  300

Qy  301  TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCTCCTGGGCCCCGGAAGAGCGTGTGG  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCTCCTGGGCCCCGGAAGAGCGTGTGG  360

Qy  361  CACGGCTCCGACCCCAGCGGGGCGGGCCGCCTGACCGACAGCTACTGCGGGCACGTGGCGGACG  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Db   361 CACGGCTCCGACCCCAGCGGCGGCGCCGCCTGACCGACCAGCAGCTACTGCGAGACGTGGCGGACG 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   361 CACGGCTCCGACCCCAGCGGCGGCGCCGCCTGACCGACCAGCAGCTACTGCGAGACGTGGCGGACG 420

Qy   421 GAGGCCCCGGCGGCCACCGGGCGAGGCGTCGTCGCTGCTGCTGGGCGGCAGGCGTGCTGGGAGCAG 480
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAGGCCCCGGCGGCCACCGGGCACCGGGCAGGCGTCGTCGCTGCTGCTGGGCGGCAGGCGTGCTGGGAGCAG 480

Db   421 GAGGCCGCGGAGTGCCGCCACCGCTTCGTGGTGCTCTGCATCGAGAAACAGCGTCATGACC 540
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   481 GAGGCCGCGGAGTGCCGCCACCGCTTCGTGGTGCTCTGCATCGAGAAACAGCGTCATGACC 540

Db   481 GAGGCCGCGGAGTGCCGCCACCGCCTTCGTGGTGCTCTGCATCGAGACAACAGCGTCATGACC 540

Qy   541 TCCTTCTCCAAG 552
            ||||||||||||
Db   541 TCCTTCTCCAAG 552

Qy   541 TCCTTCTCCAAG 552
            ||||||||||||
Db   541 TCCTTCTCCAAG 552

RESULT 4
US-10-292-418-34
; Sequence 34, Application US/10292418
; Publication No. US20030139365A1
; GENERAL INFORMATION:
;   APPLICANT: Lo, Kin-Ming
;   APPLICANT: Li, Yue
;   APPLICANT: Gillies, Stephen D
;   TITLE OF INVENTION: Expression and Export of Angiogenesis Inhibitors as
;   TITLE OF INVENTION: Immunofusins
;   NUMBER OF SEQ ID NOS: 54
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
;   LENGTH: 552
;   TYPE: DNA
;   ORGANISM: Canis familiaris
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(552)
;   OTHER INFORMATION: Endostatin
US-10-292-418-34

Query Match   99.2%;   Score 550.4; DB 14; Length 552;
Best Local Similarity 99.8%; Pred. No. 1.4e-116;
Matches 551; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CACACCCACCAGGACTTCCAGCTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 60

Qy   61 GGCGGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGTGCCGCGGCGGCCGGCGCCGGCG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGGGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGTGCCGCAGGCGGCGGCCGGCGCCGGCG 120

Qy 121 GGGCTGGCGGCCGGCACCTTCCGGCGGCCTTCCGGCGGCGCCGGCTGCGGCTGCGTCGGCGGCGGACCTCTACAGCATC 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 GGGCTGGCGGCCGGCACCTTCCGGCGGCCTTCCGGCGGCGCCGGCTGCGGCGTCGGCGGCGGACCTCTACAGCATC 180

Qy 181 GTGCGGCCGCGCCGCCGCACCGGGTGCCGTGCCCGGTGCCCCGTCAACCTCAGGGACGAGGTGCTCTTC 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 GTGCGGCCGCGCCGCCGCACCGGGTGCCGGGTGCCCGGTCAACCTCAGGGACGAGGTGCTCTTC 240

Qy 241 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAGGCCGGGCCCGGGCCCGCATC 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAGGCCCGGGCCCGCATC 300

Qy 301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCGCGGCCCCGGAAGAGCGTGTGG 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCGCGGCCCCGGAAGAGCGTGTGG 360

Qy 361 CACGGCTCCGACCCCAGCGGCGGCGCCGCCGACCAGCAGCTACTGCGAGACGTGGCGGACG 420

(second column)

Db   361 CACGGCTCCGACCCCAGCGGCGGCGCCGCCGACCAGCAGCTACTGCGAGACGTGGCGGACG 420

Qy   421 GAGGCCCCGGCGGCCACCGGGCGAGGCGTCGTCGCTGCTGGGCGGCAGGCGTGCTGGGAGCAG 480

Db   421 GAGGCCCCGGCGGCCACCGGGCAGGCGTCGTCGCTGCTGGGCGGCAGGCGTGCTGGGAGCAG 480

Qy   481 GAGGCCGCGGAGTGCCGCCACCGCTTCGTGGTGCTCTGCATCGAGAAACAGCGTCATGACC 540

Db   481 GAGGCCGCGGAGTGCCGCCACCGCCTTCGTGGTGCTCTGCATCGAGACAACAGCGTCATGACC 540

Qy   541 TCCTTCTCCAAG 552

Db   541 TCCTTCTCCAAG 552

RESULT 5
US-10-042-347-4
; Sequence 4, Application US/10042347
; Publication No. US20030114370A1
; GENERAL INFORMATION:
;   APPLICANT: O'Reilly, Michael S.
;   APPLICANT: Folkman, M. Judah
;   TITLE OF INVENTION: Nucleic Acid Molecules Encoding Endostatin Protein and Peptide F
;   TITLE OF INVENTION: Thereof
;   FILE REFERENCE: 05213-0880   (43170-249874)
;   CURRENT APPLICATION NUMBER: US/10/042,347
;   CURRENT FILING DATE: 2002-01-11
;   PRIOR APPLICATION NUMBER: US 09/315,689
;   PRIOR FILING DATE: 1999-05-20
;   PRIOR APPLICATION NUMBER: US 60/106,343
;   PRIOR FILING DATE: 1998-10-30
;   PRIOR APPLICATION NUMBER: US 09/154,302
;   PRIOR FILING DATE: 1998-09-16
;   PRIOR APPLICATION NUMBER: US 08/740,168
;   PRIOR FILING DATE: 1996-10-22
;   PRIOR APPLICATION NUMBER: US 60/005,835
;   PRIOR FILING DATE: 1995-10-23
;   PRIOR APPLICATION NUMBER: US 60/023,070
;   PRIOR FILING DATE: 1996-08-02
;   PRIOR APPLICATION NUMBER: US 60/026,263
;   PRIOR FILING DATE: 1996-09-17
;   NUMBER OF SEQ ID NOS: 6
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 546
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-042-347-4

Query Match   75.3%;   Score 418; DB 14; Length 546;
Best Local Similarity 85.3%; Pred. No. 2.1e-86;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 60
         |||||||||||||||||||||||||||||||||||||| |||||||
Db   1 CACAGCCACCGGGACTTCCAGCTTCCAGCTCCAGCTCCACCTGGTGGCTCAACAGCCCCCCTGTCA 60

Qy   61 GGCGGCATGCGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGTGCCGCGGCCGGCGCCGGCG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGCGGCATGCGAGGCATCCGGGGCATCCGCGGCGGCGACTTCCAGTGCTTCCAGTGCCGCAGGCGGCCGGCCGGTG 120

Qy 121 GGGCTGGCGGCCGGCACCTTCCGGCGGCCTTCCGGCGGCGCCGGCTGCGGCTGCGTCGGCGGCGGACCTCTACAGCATC 180
         |||||||||||||||||||||||||||||||||||
Db 121 GGGCTGGCGGCCGGCACCTTCCGGCTGCGTCGGCGGCTGCGTCGGCGGCCGGCTGCAACCTGTACAGCATC 180

Qy 181 GTGGCCCGCGCCGCCGCACCGGGTGCCCGGTGCCCCGTCAACCTCAGGGACGAGGTGCTCTTC 240
         |||||||||||||||||||||||||||||||||||||||
Db 181 GTGGCCCGTGCCGACGCGCAGCCGGCCGTGCCGCCATCGTCAACCTCAAGTGCAACTGCTGTT 240

Qy 241 CCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAGGCCCGGGCCCGCATC 300
         ||||||||||||||||||||||||||||||||||||||||||
Db 241 CCCAGCTGGGAGGCTCGTTCTTCAGGTCTGAGGGTCTGAGGGTCTCGGGCACGCATC 300

RESULT 6
US-10-131-241-53
; Sequence 53, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer
; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-131-241-53

Query Match   75.3%; Score 418; DB 14; Length 549;
Best Local Similarity   85.3%; Pred. No. 2.1e-86;
Matches 466; Conservative   0; Mismatches 80; Indels   0; Gaps   0;

```
Qy      1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG   60
            ||||  ||||||| |||||||||||||| | |||||||||||||  |||||||||| ||||
Db      1 CACAGCCACCGGGACTTCCAGCTGCGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA   60

Qy     61 GGGGGCATGCGAGGCATCCGGGAGGCGGACTTCCAGTGCTTCCAGCAGGCGCGCGCCGCG  120
            ||||| |||||  |||||||||||||||||| ||||||||||||||||||||  |||||
Db     61 GGCGGCATGCGGGGCATCCGGGGGCCGACTTCCAGTGCTTCCAGCAGGCGGCCGGGCCGTG  120

Qy    121 GGGCTGGCCGGGCACCTTCCGGGCCTGTCGCGGGCTGCAGGACCTTCTACAGCATC  180
            ||||||||||  |||||||||||||||||||||||||||||||||| |||||||||||
Db    121 GGGCTGGCGGCGCACCTTCCGGCCTCCGCGGCCTGCAGGACCTGTACAGCATC  180

Qy    181 GTGGCCGCGGCCGACCGACACCGGGGTGCCCGTCAACCTCAGGGACGAGGTGCTCTTC  240
            ||||||||||||| |||||||||||||||||| |||||||||||||||||||| |||
Db    181 GTGCCGCCGTGCCGACCGGGCCAGCCGTGCCATCGTCAACCTCAAGGACGAGCTGCTCGTT  240

Qy    241 CCCAGCTGGAGGGCCTTATTCTCGGAGGGCCAGCTGAAGCCCGGGGCCCGCATC  300
            ||||||||||||| |||| ||||||||||||||||||||||||||||| ||||||||||
Db    241 CCCAGCTGGGAGGGCTCGTTCTCAGGCTCTCAGGCTGAGGGTCCGCTGAAGCCCGGGGCACGCATC  300

Qy    301 TTCTCTTTGACGGCAGAGATGTCCTGCAGCAGTGTTGG  360
```

RESULT 7
US-10-292-418-3
; Sequence 3, Application US/10292418
; Publication No. US20030139365A1
; GENERAL INFORMATION:
; APPLICANT: Lo, Kin-Ming
; APPLICANT: Li, Yue
; APPLICANT: Gillies, Stephen D
; TITLE OF INVENTION: Expression and Export of Angiogenesis Inhibitors as
; TITLE OF INVENTION: Immunofusins
; FILE REFERENCE: LEX-006C1
; CURRENT APPLICATION NUMBER: US/10/292,418
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: 09/383,315
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: US 60/097,883
; PRIOR FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(549)
; OTHER INFORMATION: endostatin
US-10-292-418-3

Query Match   75.3%; Score 418; DB 14; Length 549;
Best Local Similarity   85.3%; Pred. No. 2.1e-86;
Matches 466; Conservative   0; Mismatches 80; Indels   0; Gaps   0;

```
Qy      1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG   60
            ||||  ||||||| |||||||||||||| | |||||||||||||  |||||||||| ||||
Db      1 CACAGCCACCACCGGGACTTCCAGCTGCGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA   60

Qy     61 GGGGGCATGCGAGGCATCCGGGAGGCGGACTTCCAGTGCTTCCAGCAGGCGCGCGCCGCG  120
            ||||| |||||  |||||||||||||||||| ||||||||||||||||||||  |||||
Db     61 GGCGGCATGCGGGGCATCCGGGGGCCGACTTCCAGTGCTTCCAGCAGGCGGCCGGGCCGTG  120

Qy    121 GGGCTGGCCGGGCACCTTCCGGGCCTGTCGCGGGCTGCAGGACCTTCTACAGCATC  180
            ||||||||||  |||||||||||||||||||||||||||||||||| |||||||||||
Db    121 GGGCTGGCGGCGCACCTTCCGGCCTCCGCGGCCTGCAGGACCTGTACAGCATC  180

Qy    181 GTGGCCGCGGCCGACCGACACCGGGGTGCCCGTCAACCTCAGGGACGAGGTGCTCTTC  240
            ||||||||||||| |||||||||||||||||| |||||||||||||||||||| |||
Db    181 GTGCCGCCGTGCCGACCGGGCCAGCCGTGCCATCGTCAACCTCAAGGACGAGCTGCTCGTT  240

Qy    241 CCCAGCTGGAGGGCCTTATTCTCGGAGGGCCAGCTGAAGCCCGGGGCCCGATC  300
            ||||||||||||| |||| ||||||||||||||||||||||||||||| ||||||||||
Db    241 CCCAGCTGGGAGGGCTCGTTCTCAGGCTCTCAGGCTGAGGGTCCGCTGAAGCCCGGGGCACGCATC  300
```

```
Qy    301 TTCTCTTTGACGGCAGAGATGTCCTGAGCACCCCGCCCCCGGAAGAGCGTGTGG 360
          ||||||||||||||||||||| ||||||||| ||||||||||||||||||||||
Db    301 TTCTCCTTTGACGGCAAGGACGTCCTGAGGCACCCACCTGGCGCTGGTGTGG 360

Qy    361 CACGGCTCCGACCCCAGCCGGCGGCGGCGCCTGACCAGCACTGTGGAGACGGCGGACG 420
          ||||||||||||||||||||| ||||||||||||||||||||||||| ||||||||
Db    361 CATGGCTCCGACCCCAAGGGCGCAGCGTGGCCTGACGTGCTGCTGGCGGACG 420

Qy    421 GAGGCCCGCGGCGACTGCCCGCCGCTGCGGTGCGGGCCAGCGTGCTGCTGAGCAG 480
          |||||||||||| ||||||||| |||||||||||||||||||||||||||||
Db    421 GAGGCTCCCTGCGCCCAGGCGCCCAGGCCTCCTGGGGCAG 480

Qy    481 GAGGCCGCGAGTGTCCGCCACGCCTTCGTCGTCTTCGTCGTGGTGGCTGAGCAG 540
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGTGCCGGAGCTGCCATCACGCGCTACATCGTGCTCTTGCATTGAGAACAGCAGCTTCATGACT 540

Qy    541 TCCTTC 546
          ||||
Db    541 GCCTCC 546


RESULT 8
US-09-873-676-30
; Sequence 30, Application US/09873676
; Patent No. US2002007289A1
; GENERAL INFORMATION:
; APPLICANT: MacDonald, Nicholas J.
; APPLICANT: Sim, Kim L.
; TITLE OF INVENTION: Angiostatin and Endostatin Binding Proteins and Methods of Use
; FILE REFERENCE: 05213-0378 (43170-259333)
; CURRENT APPLICATION NUMBER: US/09/873,676
; CURRENT FILING DATE:  2001-06-04
; PRIOR APPLICATION NUMBER: US 60/209,065
; PRIOR FILING DATE:  2000-06-02
; PRIOR APPLICATION NUMBER: US 60/289,387
; PRIOR FILING DATE:  2001-05-08
; SOFTWARE: PatentIn version 3.1
; NUMBER OF SEQ ID NOS: 123
; SEQ ID NO 30
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-873-676-30

Query Match       75.3%;  Score 418;  DB 9;  Length 552;
Best Local Similarity  85.3%;  Pred. No. 2.1e-86;
Matches 466;  Conservative   0;  Mismatches   0;  Indels   80;  Gaps   0;

Qy      1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 60
          || |||||||||||||||||||||||||||||||||||| |||||||||||||
Db      1 CACAGCCACCGGACTTCCAGCCGGTGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA 60

Qy     61 GGCGGCATGCGAGGCATCCGGGGGCGGGGCCTTCCGCGTCGCGGGGCCGGGGCCCGTG 120
          ||||||||||||||||||||||||||||||||||||| ||||| ||||||||||||
Db     61 GGCGGCATGCGGGGGCATCCGCGGGGCGGGGCCGACTTCCAGCTGCTTCCAGCGTCGCGGGACCGGGGCCCGTG 120

Qy    121 GGGCTGCGCCGGCGCACCTTCCGGGCCTTCCTGTCGTCGTCGCGGGCCTCCTGAGGCTGCGCCGCCTTCTCGCGGCTTCCTGCGCTGCGCCTCTACAGCATC 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGGCTGCGCCGGCGCACCTTCCGCGGCCTTCCTGTCGTCGTCGCGGCCGCCTGCGGCCTGCGACCTGTACAGCATC 180

Qy    181 GTGCGGCGGCGCCGACCGGCCACGGGGTGCCCGTGCTCAACCTCAGGGACGAGGTGCTCTTC 240
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db    181 GTGCGGCCGTGCCGACGCCGCAGCCGCAGCCGGCCCATCGTCAACCTCAACCTCAAGGACGACGAGCTGCTGTTT 240

Qy    241 CCAGCTGGGAGGGCCTTATTCTCGGGCTCCAGGGGGCCCGGCATC 300
          ||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCCAGCTGGGAGGGCTCGGCGTCTGCAGCAGCCTCGAGCGCTCGAGGGCCACGCGGGCACGCGCATC 300

Qy    301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCCCGGCGCCCGGAAGAGCGTGTGG 360
          ||||||||||||||||||||| ||||||||| ||||||||||||||||||||||
Db    301 TTCTCCTTTGACGGCAAGGACGTCCTGAGGCACCCACCTGGCGCTGGTGTGG 360

Qy    361 CACGGCTCCGACCCCAGCCGGCGGCGCCGCCTGACCAGCACTGTGGAGACGGCGGACG 420
```

```
RESULT 9
US-09-880-107-2178
; Sequence 2178, Application US/09880107
; Patent No. US2002014298lAl
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE:  2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE:  2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE:  2000-10-02
; SOFTWARE: PatentIn Ver. 2.1
; NUMBER OF SEQ ID NOS: 3950
; SEQ ID NO 2178
; LENGTH: 3394
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US2002014298lAl L22548
US-09-880-107-2178

Query Match       75.3%;  Score 418;  DB 9;  Length 3394;
Best Local Similarity  85.3%;  Pred. No. 1.6e-86;
Matches 466;  Conservative   0;  Mismatches   0;  Indels   80;  Gaps   0;

Qy      1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 60
          || ||||||||||||||||||||||||||||||||||| |||||||||||||
Db   1504 CACAGCCACCGGACTTCCAGCCGGTGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA 1563

Qy     61 GGCGGCATGCGAGGCATCCGGGGGCGGGGCCTTCCGCGTCGCGGGGCCGGGGCCCGTG 120
          |||||||||||||||||||||||||||||||| |||||| ||||| ||||||||||||
Db   1564 GGCGGCATGCGGGGGCATCCGCGGGGCGGGGCCGACTTCCAGCTGCTTCCAGCGTCGCGGGACCGGGGCCCGTG 1623

Qy    121 GGGCTGCGCCGGCGCACCTTCCGGGCCTTCCTGTCGTCGTCGCGGGCCTCCTGAGGCTGCGCCGCCTTCTACAGCATC 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1624 GGGCTGCGCCGGCGCACCTTCCGCGGCCTTCCTGTCGTCGTCGCGGCCGCCTGCGACCTGTACAGCATC 1683

Qy    181 GTGCGGCGGCGCCGACCGGCCACGGGGTGCCCGTGCTCAACCTCAGGGACGAGGTGCTCTTC 240
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   1684 GTGCGGCCGTGCCGACGCCGCAGCCGCAGCCGGCCCATCGTCAACCTCAACCTCAAGGACGACGAGCTGCTGTTT 1743

Qy    241 CCAGCTGGGAGGGCCTTATTCTCGGGCTCCAGGGGGCCCGGCATC 300
          ||||||||||||||||||||||||||||||||||||||||||
Db   1744 CCCAGCTGGGAGGGCTCGGCGTCTGCAGCAGCCTCGAGCGCTCGAGGGCCACGCGGGCACGCGCATC 1803

Qy    301 TTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGCCCGGCGCCCGGAAGAGCGTGTGG 360
          ||||||||||||||||||||| ||||||||| ||||||||||||||||||||||
Db   1804 TTCTCCTTTGACGGCAAGGACGTCCTGAGGCACCCACCTGGCACCCCACCTGGCGCTGGTGTGG 1863

Qy    361 CACGGCTCCGACCCCAGCCGGCGGCGCCGCCTGACCAGCACTGTGCGAGACGTCGGCGGACG 420
```

Db   4420 GAGGCTCCCTGCGGCCACAGGGCGCCAGGCTTCACCGAGAGCTACTGTGAGACGTGGCGGACG 4479
Db   1864 CATGGCTCGGACCCCACGGGCGCAGGCTACTGTGAGACGTGGCGGACG 1923

Qy   481 GAGGCCGGAGCTGCGCGCCACGGCCTTCGTGGTGCTCTGCGGGCGAGGCTGCTGGAGCAG 540
Qy   421 GAGGCCCGGGCGGCCACCGGGCCAGGCGTGTCTGCTGCGGGCGAGGCTGCTGGAGCAG 480

Db   4480 AGTGCGCCGGAGCTGCCATCACGCGCCTACATCGTCGTGCTGCTGCATTGAGAACAGCTTCATGACT 4539
Db   1924 GAGGTCTCCCTGCGGCCACGGGCCAGGCCTCCTGGGGGCAGGCTCCTGGGGACG 1983

Qy   541 TCCTTC 546
Qy   481 GAGGCCGAGCTGCGCCACGCCTTCGTGGTGCTCTGCGGGCGAGGCTGCTGGAGCAG 540

Db   4540 GCCTCC 4545
Db   1984 AGTGCCGAGCTGCCATCACGCGCCTACATCGTCGTGCTGCTGCATTGAGAACAGCTTCATGACT 2043

Qy   541 TCCTTC 546

Db   2044 GCCTCC 2049

RESULT 11
US-10-264-049-835
; Sequence 835, Application US/10264049
; Publication No. US2004000579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: Patentin Ver. 3.1
; SEQ ID NO 835
; LENGTH: 4875
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-264-049-835

Query Match   75.3%;   Score 418;   DB 15;   Length 4875;
Best Local Similarity 85.3%;   Pred. No. 1.5e-86;
Matches 466;   Conservative 0;   Mismatches 80;   Indels 0;   Gaps 0;

Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 60
Db   2512 CACAGCCACCGCGACTTCCAGCTGGTGCTCCAACCTCGTTGCGCTCAACAGCCCCCTGTCA 2571

Qy   61 GGGGGCATGCGAGGCATCGGGGAGCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGCGGCGGCCGCG 120
Db   2572 GCGGGCATGCGAGGCATCGGGGGCACCTTCCGGCGCCTCCGCGCGGCCGGTG 2631

Qy   121 GGGCTGGCCGGCCATCGGGGAGCCGGACTTCCAGTGCTTCCGCGTGTCCCGTCGTCAACCTCGGGGCG 180
Db   2632 GGGCTGGCGCGGGAGCCGGCACCTTCCGGCGCCTCCGCGCGGCCGGTG 2691

Qy   181 GTGCGCGCCGGCCACCGGGCTGCCGGTGCCGGTCGTCAACCTTCAGGGACGAGGTGCTCTTC 240
Db   2692 GTGCGCCGGCGCCGGCCATCGTGCCATCGTCAACCTCAAGGACGAGCTGTGTTT 2751

Qy   241 CCCAGCTGGGAGGCCTTATTCTCGGGGCTCGAGGGCCAGCAGCCCCCATC 300
Db   2752 CCCAGCTGGGAGGCTCTGTTCTCAGGCTCTGAGGGCACCCCACCTGAAGCCCACGCAGCATC 2811

Qy   301 TTCTCTTTGACGGCGGAGATGTCCTGCAGCAGAGGACGAGGTGCTCTTC 360
Db   2812 TTCTCCTTTGACGGCAGGCAAGGACGCTCCTGAGGCACACCCACCTGCGAGACGTGTGG 2871

Qy   361 CACGGCTCCGAGCTGCGGACCCCAGGCGGGCCGGCCAGCGGTCGTGCTGCTGCCGGGGCG 420
Db   2872 CATGGCTCGGACCCCACGGGCGCAGGCTACTGTGAGACGTGGCGGACG 2931

Qy   421 GAGGCCCGGGCGGCCACCGGGCCAGGCGTGTCTGCTGCGGGCGAGGCTGCTGGAGCAG 480
Db   2932 GAGGTCTCCCTGCGGCCACGGGCCAGGCTCCTGGGTGGTGCTCTGCGGGCGAGGCTCATGACC 2991

Qy   481 GAGGCCGAGCTGCGCCGCCACGCCTTCGTGGTGCTCTGCGGTGCTCCATCGATCGAGAACAGCTTCATGACC 540
Db   2992 AGTGCGCCGGAGCTGCCATCACGCGCTACATCGTCGTGCTGCTGCATTGAGAAACAGCTTCATGACT 3051

```
Qy      541  TCCTTC  546
             ||||||
Db      3052  GCCTCC  3057
```

RESULT 12
US-10-080-797-2
; Sequence 2, Application US/10080797
; Publication No. US2002018253A1
; GENERAL INFORMATION:
;   APPLICANT: Campochiaro, Peter A.
;   APPLICANT: Dixon, Katharine H.
;   APPLICANT: Brazzell, Romulus K.
;   TITLE OF INVENTION: METHOD FOR TREATING OCULAR
;   TITLE OF INVENTION: NEOVASCULARIZATION
;   FILE REFERENCE: 4-31881A
;   CURRENT APPLICATION NUMBER: US/10/080,797
;   CURRENT FILING DATE: 2002-02-21
;   NUMBER OF SEQ ID NOS: 21
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 551
;   TYPE: DNA
;   ORGANISM: Human
US-10-080-797-2

Query Match   75.1%; Score 417; DB 13; Length 551;
Best Local Similarity 85.3%; Pred. No. 3.5e-86;
Matches 465; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy    2  ACACCCACCAGGACTTCCAGTCGGTGCTGCCCTGAACAGCCCGCAGCCGG   61
         ||| |||||||||  || ||||||||||||||||||||||||||||| |||
Db    1  ACAGGCACCGGACTTCCAGCCGGTGCTCCACCTGGTTGCGCTCAACAGCCCCGTCTCAG   60

Qy   62  GCGGGCATGCGAGGCATCCGGGGAGGCAGGACTTCCAGCAGGCGGCGCGCCGGCCGCGCCGG  121
         |||||||| |||||| ||||||||||||||||||||||||||||||||||||
Db   61  GCGGGCATGCGGGGGCATCCGGGGGCCGGGGGCATCCCCAGTGCTTCCAGCGTGCTTCCAGCGTGCTTCCAGCGGGGCCGTGG   120

Qy  122  GGCTGGCCGCGCCACCTTCCTGTCGTCGCCGCTGCGAGGACCTTCTACAGCATCG   181
         ||||||||||||||||||||||||||||||||||||||
Db  121  GGCTGGCGGGCACCTTCCCGCGCCTTCCTGTCCTCGCCGCTGCGAGGACCTGTACAGCATCG   180

Qy  182  TGCGGCCGCGCCGACCGCACCGGGGTGCCCGTGCCATCGGGGACGACGAGGTGCTCCTCC   241
         ||||||||||||||||||||||||||||||||||||||
Db  181  TGCGCGGCCGTGCCGACCGGCAGCCGTGCCCATCGTGCCATCGTCAAGGACGACGTGCTGTTTC   240

Qy  242  CCAGCTGCTGGGGAGGCCTTATTCTCGGGCTCCGAGGCCCAGCTGGGACGCACCCCGCCATCT   301
         |||||||||||||||||||||||||||||||||||||||||
Db  241  CCAGCTGGGGAGGCCTCTGTTCTCAGGCTCTGAGGTCGCTCGAGGGTCTGAAGTCCGCTGAAGCGCACGCACCATCT   300

Qy  302  TCTCTTTGACGCGCAGAGATGTCCTGCAGCACCCGCCTGCCCGCCACCCGCCGTGTGGC   361
         ||||||||||||||||||||||||||||||||||
Db  301  TCTCCTTTGACGCGCAAGGACGTCCTGAGGCAAGTGCTGGCCCAACTCCTGGCCCCACCTGGCCCGCCACCTGGCCGTGTGGC   360

Qy  362  ACGGCTCCGACCCCAGCGGGCGCGCCCAGGCGGGCGCCACCGGGGGCCAC   421
         ||||||||||||||||||||||||||||||||||||
Db  361  ATGGCTTCGGACCCCAACGGGCGCCAGGCGCGCCGCGCGAGACGCTGGGCGGGACGG   420

Qy  422  AGGCCCCGCCGCCACCGGGGCCCACCCGAGGGCGCGCCTCCTCGGGGGGCCCAGG   481
         |||||||||||||||||||||||||||||
Db  421  AGGCTCCCTCGCGGGCCACGCGCCCACCCGAGGGCCAGGCTCCTCGCTGCTGGGGGCCAGA   480

Qy  482  AGGCCGGGGGGAGCTGCGCCGCCCGAGCTGCGCCATCGAGAACAGCGTCATGACCT   541
         ||||||||||||||||||||||||||||
Db  481  GTGCCGGGGGGAGCTGCGCCGCCGCCGCCATCACGCGCCTACATCGTGCTCTGCATTGAGAACATCATGACTG   540

Qy  542  CCTTC  546
         |||||
Db  541  CCTCC  545
```

RESULT 13
US-10-131-241-51

; Sequence 51, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
;   APPLICANT: Holaday, John W.
;   APPLICANT: Fortier, Anne H.
;   TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolife
;   TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
;   FILE REFERENCE: 05213-0344   43170-271565
;   CURRENT APPLICATION NUMBER: US/10/131,241
;   CURRENT FILING DATE: 2002-07-22
;   PRIOR APPLICATION NUMBER: US 09/413,049
;   PRIOR FILING DATE: 1999-10-06
;   PRIOR APPLICATION NUMBER: US 09/316,802
;   PRIOR FILING DATE: 1999-05-21
;   PRIOR APPLICATION NUMBER: US 60/086,586
;   PRIOR FILING DATE: 1998-05-22
;   NUMBER OF SEQ ID NOS: 65
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51
;   LENGTH: 632
;   TYPE: DNA
;   ORGANISM: Murinae sp.
US-10-131-241-51

Query Match   74.9%; Score 415.6; DB 14; Length 632;
Best Local Similarity 82.8%; Pred. No. 7.2e-86;
Matches 523; Conservative 0; Mismatches 29; Indels 0; Gaps 1;

```
Qy    1  CACACCCACCAGGACTTCCAGTCGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG   60
         |||| ||| || ||||  || ||||||| ||||||||||||
Db    1  CATACTCATCAGGACTTTCAGCAGTGCTTCCACCTGGTGGCACTGAACACCCCCGTCT   60

Qy   61  GGCGGCATGCGAGGCATCCGGGGAGGCAGGACTTCCAGTCGTTCCAGTGCTTCGGCGCGCCGCG   120
         ||||||||| ||||| ||||||||||||||||||| ||||||||||||
Db   61  GGAGGCATGCGTGGGTATCCGTGGGCAGCGAGATTTCCAGTGTCGCAAGCCCGAGCCGTG   120

Qy  121  GGGCTGGCGCGGCCGCACCTTCCTGTCGTCGGGC-----------   160
         |||| | ||| |||||| ||||  ||||||| |||
Db  121  GGGCTGTCGGGCACCTTCGGGGCTTCCGGTCCCTCAGGCTGCGAGGATCTCTATAGCATC   180

Qy  161  ----------------   160
Db  161  GTGCGGCCGTGCTGACCGGGGGTCTGTGCCCATCGTGCCACGGGTGCCCGTCGTCAACC   240

Qy  181  GTGCGGCCGTGCTGACCGGGGGTCTGTGCCCATCGTCAACCTGAAGGACGAGGTGCTATCT   240
         ||||||||||||||| | |||||||||||||||||||||
Db  181  TGCAGGACCTCTACAGCATCGTGCGCGCCGACCGCACCGGGGTGCCCGTCGTCAACC   220

Qy  241  TGCAGGACCTCTACAGCATCGTGCGCGCCGACCGCACCGGGGTGCCCGTCGTCAACC   300
Db  241  TGCAGGACCTCTACAGCATCGTGCGCGCCGACCGCACCGGGGTGCCCGTCGTCAACC   300

Qy  221  TCAGGGACGAGGTGCTCTTCCCCAGCTGGGAGGCCTCTCGGGGCTCCGAGGGCCAGC   280
         ||||||||||||||||| | ||||||||||||||||||||||
Db  301  TCAGGGACGAGGTGCTCTTTATTCTCGGGCTCCGAGGGCCAGC   360

Qy  301  TCAGGGACGAGGTGCTCTTTATTCTCGGGCTCGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGC   360
Db     

Qy  281  TGAAGCCCGGGCCCGCCATCTTCTTTCGACCGCAGAGATGTCCTGCAGCACCCGCCT   340
         |||||||||||||||| |||||||||||||||||||||
Db  281  TGAAGCCCGGGCCCGCCATCTTCTTTCTTTCGACCGCAGAGATGTCCTGCAGCACCCGCCT   340

Qy  361  GGCCCCGGAAGAGCGTGTGCACGGCGTCCGACCGCACCCCAGCGGGGCGGCCTGACCGACAGCT   400
Db  361  TGAAGCCCGGGCCCGCCATCTTCTTTCTTTCGACCGCAGAGATGTCCTGCAGCACCCGCCT   420

Qy  341  GGCCCCGGAAGAGCGTGTGCACGGCGTCCGACCGCACCCCAGCGGGGCGGCCTGACCGACAGCT   400
Db     

Qy  421  GGCCCCGGAAGAGCGTGTGCGACGAGGCCCAGGCCCCGCCGGCGTCGTGCTGGG   480
Db  421  GGCCCCGGAAGAGCGTGTGCGACGAGGCCCACCGCCCGGCGGCGTCGTGCTGGG   480

Qy  401  ACTGCGAGACGTGCTGGACGAGCAGGAGGCGGAGCTGCGTCGTGG   460
Db  481  ACTGCGAGACGTGGCGACGTGGACGAGCAGGAGCGCCGGAGCTGCGTCGTGG   540

Qy  461  CGGGCAGGCGTCGTGAGCAGCAGGAGGCGGAGCTCCGCGACGCTTCGTCGTGCTCTGCA   520
         |||||||||||||||||||||||||||||||||||||
Db  541  CGGGGCAGGCGTCTGAGCAGCAGGAGGAGGCGGAGCTGCGCGACGCCGCTTCGTCGTGCTCTGCA   600

Qy  521  TCGAACACGCTCATGACCTCCTTCTCCAAG  552
         ||||| ||||||| |||||||||||||||
Db  601  TCGAGAACACGCTCATGACCTCCTTCTCCAAG  632
```

RESULT 14
US-10-210-172-161
; Sequence 161, Application US/10210172
; Publication No. US20040043928A1
; GENERAL INFORMATION:
;   APPLICANT: Kekuda, Ramesh
;   APPLICANT: Miller, Charles
;   APPLICANT: Patturajan, Meera
;   APPLICANT: Pena, Carol
;   APPLICANT: Rieger, Daniel
;   APPLICANT: Shimkets, Richard
;   APPLICANT: Zerhusen, Bryan
;   APPLICANT: Li, Li
;   APPLICANT: Ji, Weizhen
;   APPLICANT: Padigaru, Muralidhara
;   APPLICANT: Casman, Stacie
;   APPLICANT: Voss, Edward
;   APPLICANT: Boldog, Ferenc
;   APPLICANT: Gorman, Linda
;   APPLICANT: Leite, Mario
;   APPLICANT: Vernet, Corine
;   APPLICANT: Anderson, David
;   APPLICANT: Guo, Xiaojia
;   APPLICANT: Zhong, Mei
;   APPLICANT: Gerlach, Valerie
;   APPLICANT: Hjalt, Tord
;   APPLICANT: Rastelli, Luca
;   APPLICANT: Spytek, Kimberly
;   APPLICANT: Edinger, Shlomit
;   APPLICANT: Ellerman, Karen
;   APPLICANT: Malyankar, Uriel
;   APPLICANT: MacDougall, John
;   APPLICANT: Stone, David
;   APPLICANT: Alsobrook II, John
;   APPLICANT: Lepley, Denise et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-416 A
; CURRENT APPLICATION NUMBER: US/10/210,172
; CURRENT FILING DATE: 2001-08-01
;   PRIOR APPLICATION NUMBER: 60/309,501
;   PRIOR FILING DATE: 2001-08-02
;   PRIOR APPLICATION NUMBER: 60/323,994
;   PRIOR FILING DATE: 2001-09-21
;   PRIOR APPLICATION NUMBER: 60/373,814
;   PRIOR FILING DATE: 2002-04-19
;   PRIOR APPLICATION NUMBER: 60/310,291
;   PRIOR FILING DATE: 2001-08-03
;   PRIOR APPLICATION NUMBER: 60/310,951
;   PRIOR FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: 60/310,544
;   PRIOR FILING DATE: 2001-08-07
;   PRIOR APPLICATION NUMBER: 60/311,292
;   PRIOR FILING DATE: 2001-08-09
;   PRIOR APPLICATION NUMBER: 60/311,979
;   PRIOR FILING DATE: 2001-08-13
;   PRIOR APPLICATION NUMBER: 60/313,201
;   PRIOR FILING DATE: 2001-08-17
;   PRIOR APPLICATION NUMBER: 60/312,892
;   PRIOR FILING DATE: 2001-08-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 327
; SOFTWARE: CuraSeqList version 0.1
;
; SEQ ID NO 161
;   LENGTH: 555
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(552)
US-10-210-172-161

Query Match   74.2%;   Score 411.6;   DB 12;   Length 555;
Best Local Similarity   84.6%;   Pred. No. 66-85;

Matches 462;   Conservative   0;   Mismatches   84;   Indels   0;   Gaps   0;

Qy   1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGGCCCTGACCAGCCCGCAGCCG 60
            |||||| || ||| ||||||||||| |||||||| ||| ||| ||| || ||||||
Db   4 CACAGCCACCCGCACTTCCAGCCCGGGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA 63

Qy   61 GGGGCATGCGAGGCATCCGGGAGGCGGACTTCCAGTGTCTTCCAGCAGGCGGCGGCGCCG 120
            ||||||||| ||||||||| ||| |||||||||||||| ||||||||||||||||||
Db   64 GGCGGGCATGCGGGGGCATCCGGGGCCGACTTCCAGTGTCTTCCAGCAGCGCGGGGCCGTG 123

Qy   121 GGGCTGCCCGGCACCTTCCGGGCCTTCCTGTCCGTCGCGGCTGCAGCACCTCTACAGCATC 180
            |||||||||||||||||||||| |||||| ||||| ||||||| ||||||||||||||
Db   124 GGGCTGCGGGGCACCTTCCGCGCCCTTCCGTGCTCGCGGCTGCAGGACCTGTACAGCATC 183

Qy   181 GTGCGGCGCGCCGACCGGCACCGGGGTGCCCGTCGCGGTGCACCTCAACCTCAGGGACGAGC 240
            |||||||| ||||||||||||| |||| ||||||||||||| |||||||||||||||
Db   184 GTGCGGCCGTGCCGACCGCGACCGGCAGCCGTGCCCATCGTGCCCATCTCAACCTCAAGGACGAGCTGCTGTTT 243

Qy   241 CCCAGCTGCGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCCGCATC 300
            ||||||||| ||||||| |||| |||||||| |||||||||||||||||||||||||||
Db   244 CCCAGCTGGGGAGGCTCTGTTCTCAGGCTCTGAGGGTCCGGTCAAGCCCGGGGCACGCATC 303

Qy   301 TTCTCTTTCGACGGCGAGAGATGCTCTGCAGCACCCCGCTCGGCCCTCGGCCCCGAAGAGAGCGTGTGG 360
            ||||||||| |||| |||| ||||||||| ||||||||||||||||| |||||||||||
Db   304 TTCTCCTTTAACGGCAAGGACGTCCTGACCCACCCACCTGGCCTCGGCCCTCGGCCCCAGAAAGAGCGTGTGG 363

Qy   361 CACGGCTCCGACCCCAGCGGCGCGCCTGACCGACCAGCTACTGCGAGACGTGGCCGGACG 420
            |||||||||||||||||||||||||||| |||||| |||||||||||||||||||||||
Db   364 CATGGCTCGGACCCCAACGGCGCAGGCTGACAGCGAGCTACTGTGAGACGTGGCCGGACG 423

Qy   421 GAGGCCCCGGCGGGCCACCCGGGCAGGCGTCGTCGCTGCTGGCGGGCCAGGCTGCTGGAGCAG 480
            |||||||| ||||||||||||||||||||||||||||| |||||||||||||||||||
Db   424 GAGGCTCCCTCCGGCCACCAGGGCCAGGCCTACTGCGTGCTGTCGGGGGCAGGCTCCTGGGGCAG 483

Qy   481 GAGGCCGGAGCTGCCGCCACGCCTTGGTGGTGCTGCTCGATCGAGAACAGCGTCATGACC 540
            ||||| || ||||| ||||| |||| ||||| ||| | || |||||||| ||||||
Db   484 AGTGCCGGAGCTGCCATCACGCCTCACATCGTGCTATGCGTCGTATGCATTGAGAACAGCTTCATGACT 543

Qy   541 TCCTTC 546
            ||||||
Db   544 GCCTCC 549

RESULT 15
US-10-042-347-6
; Sequence 6, Application US/10042347
; Publication No. US2003011437OA1
; GENERAL INFORMATION:
;   APPLICANT: O'Reilly, Michael S.
;   APPLICANT: Folkman, M. Judah
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Endostatin Protein and Peptide |
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 05213-0880   (43170-249874)
; CURRENT APPLICATION NUMBER: US/10/042,347
; CURRENT FILING DATE: 2002-01-11
;   PRIOR APPLICATION NUMBER: US 09/315,689
;   PRIOR FILING DATE: 1999-05-20
;   PRIOR APPLICATION NUMBER: US 60/106,343
;   PRIOR FILING DATE: 1998-10-30
;   PRIOR APPLICATION NUMBER: US 09/154,302
;   PRIOR FILING DATE: 1998-09-16
;   PRIOR APPLICATION NUMBER: US 08/740,168
;   PRIOR FILING DATE: 1996-10-22
;   PRIOR APPLICATION NUMBER: US 60/005,835
;   PRIOR FILING DATE: 1995-10-23
;   PRIOR APPLICATION NUMBER: US 60/023,070
;   PRIOR FILING DATE: 1996-08-02
;   PRIOR APPLICATION NUMBER: US 60/026,263
;   PRIOR FILING DATE: 1996-09-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin version 3.1
;
; SEQ ID NO 6
;   LENGTH: 534

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-042-347-6

Query Match          74.0%;   Score 410.8;  DB 14;  Length 534;
Best Local Similarity 85.6%;   Pred. No. 9.2e-85;
Matches 457;  Conservative 0;  Mismatches 77;  Indels 0;  Gaps 0;

Qy   13 GACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCGGCGGCATGCGA   72
         |||||||||||||  |||||||   |||||||||| |  |||||||| |||||||||||
Db    1 GACTTCCAGCCGGTGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCAGGCGGCATGCGG   60

Qy   73 GGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGCGCGCCGCGGGGCTGGCCGGC  132
         ||||||||||||||||  |||||||||||||||   ||| |||||||| | |||||||
Db   61 GGCATCCGCGCGGGCCGGACTTCCAGTGCTTCCAGCAGCGCCGGGGCGCCGTGGCGGGC  120

Qy  133 ACCTTCCGGGCCTTCCTGTCGTCGGCGTCGAGGACCTCTACAGCATCGTGCCCGCGCC  192
         |||||||||  ||||||||  |||||||||||||||||||||||||||||||||| |||
Db  121 ACCTTCCCGCGCCTTCCTGTCCTCGCGCGCGTCGAGGACCTGTACAGCATCGTGCCCGTGCC  180

Qy  193 GACCGCACCGGGTGCCCGTGTCAACCTCAGGGACGAGGTGCTCTTCCCCAGCTGGGAG  252
         |||||  |  ||||| |||||||||||||||||||||||||||||||||||||||||||
Db  181 GACCGCGCAGCCGTGCCCATCGTCAACCTCAAGGACGAGCTGCTGTTTCCCAGCTGGGAG  240

Qy  253 GCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGGCCCGGGGCCGCATCTTCTCTTTCGAC  312
         ||  ||  |||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db  241 GCTCTGTTTCTTCAGGCTCTGAGGGTCCGCTGAAGGCCCGGGGCACGCATCTTCTCCTTTGAC  300

Qy  313 GGCAGAGATGTCCTGCAGCACCCCGCCTGGCCCGGAAGAGCGTGTGGCACGGCTCCGAC  372
         |||||  |||||| ||||||||  ||||| |||||||||||||||||||||||||||||
Db  301 GGCAAGGACGTCCTGAGGCACCCCACCCTGGCCCCAGAGAGCGTGTGGCATGGCTCGGAC  360

Qy  373 CCCAGCGGGGCGCCCGCCTGACCAGCTACTGCGAGACGTGCGGAGGCCCCGGGCG  432
         |||||||||||||  ||||||||||||||||||||  ||||| |||||||||| |
Db  361 CCCAACGGGGCCAGCGCAGGCTGCAGGCTACGAGAGCTACTGTGAGACGTGCGGACGGGAGGCTCCCTCG  420

Qy  433 GCCACCGGGCAGGCGTCGTCGTGCTGCTGGCCGGCGCAGGAGGCCGCGAGC  492
         |||||||  |  ||||||||||||||||| |||||||||||||||||||||
Db  421 GCCACGGGCCAGGCCTGGCTGCTGCGTGCTGGGGGCGGCTCCCGGGGGCCAGAGTCCGGGAGC  480

Qy  493 TGCCGCCACGCCTTCGTGGTGCTCGCATCGAGGAACAGCCGTCATGACCTCCTTC  546
         |||  |||  |  |||||||||||||||||| |||||||||||||| ||||||||
Db  481 TGCCATCACGGCTACATCGTCCTCGCATTGAGGATTGAGAACAGCTTCATTGACTGCCCTCC  534
```

Search completed: March 29, 2004, 16:40:38
Job time : 231.15 secs

```
          GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        March 29, 2004, 11:07:01 ; Search time 1704.7 Seconds
                                          (without alignments)
                                          9722.244 Million cell updates/sec

Title:         US-09-938-391-3
Perfect score: 555
Sequence:      1 cacaccaccaggagacttcca.........tgacctccttctccaagtag 555

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      2751289 segs, 1493109276 residues

Total number of hits satisfying chosen parameters:        55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     EST:*
               1:  em_estba:*
               2:  em_esthum:*
               3:  em_estin:*
               4:  em_estmu:*
               5:  em_estov:*
               6:  em_estpl:*
               7:  em_estro:*
               8:  em_htc:*
               9:  gb_est1:*
               10: gb_est2:*
               11: gb_htc:*
               12: gb_est3:*
               13: gb_est4:*
               14: gb_est5:*
               15: em_estfun:*
               16: em_estom:*
               17: em_gss_hum:*
               18: em_gss_inv:*
               19: em_gss_pln:*
               20: em_gss_vrt:*
               21: em_gss_fun:*
               22: em_gss_mam:*
               23: em_gss_mus:*
               24: em_gss_pro:*
               25: em_gss_rod:*
               26: em_gss_phg:*
               27: em_gss_vrl:*
               28: gb_gss1:*
               29: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	439.8	79.2	551	10	BF074459	BF074459 221883 MA
2	418	75.3	707	10	BE908201	BE908201 601500458
3	418	75.3	757	10	BE906253	BE906253 601502237
4	418	75.3	881	14	CD105862	CD105862 AGENCOURT
5	418	75.3	4230	11	BC063833	BC063833 Homo sapi
6	410.8	74.0	504	14	CF789984	CF789984 867322 MA
7	394.2	71.0	944	13	BU859398	BU859398 AGENCOURT
8	391.8	70.6	657	14	CB444165	CB444165 695295 MA
9	384.4	69.3	703	13	BU615520	BU615520 UI-H-FG0-
10	384.2	69.2	929	13	BQ672290	BQ672290 AGENCOURT
11	377	67.9	683	12	BM683067	BM683067 UI-E-BO1-
12	376.6	67.9	682	9	AW089583	AW089583 xd20f03.x
13	375.8	67.7	843	10	BF385854	BF385854 60204602
14	375.8	67.7	874	12	BI412588	BI412588 602990468
15	375.8	67.7	4192	12	BC062931	BC062931 Mus muscu
16	374.8	67.5	715	14	CB596713	CB596713 AGENCOURT
17	373.2	67.2	735	13	CF728236	CF728236 UI-M-HB0-
18	371.6	67.0	832	12	BG387051	BG387051 60245474
19	370.2	66.7	979	13	BQ673186	BQ673186 AGENCOURT
20	369.6	66.6	897	12	BI080524	BI080524 60287700
21	368.6	66.4	611	10	AW911243	AW911243 ur83h10.y
22	367.2	66.2	884	12	BI161007	BI161007 602865213
23	366.6	66.1	1093	13	BQ723254	BQ723254 AGENCOURT
24	364.2	65.6	723	12	BI247582	BI247582 602960041
25	363.8	65.5	753	12	BI904605	BI904605 603168411
26	362.6	65.3	782	9	AI326391	AI326391 mm18f09.x
27	360.2	64.9	668	14	CF724654	CF724654 UI-M-GZ0-
28	357.4	64.4	618	9	AV696242	AV696242
29	350.6	63.2	664	13	BU632049	BU632049 UI-H-FE1-
30	350.6	63.2	715	9	AU125614	AU125614
31	350.6	63.2	947	13	BU556872	BU556872 AGENCOURT
32	348.6	62.8	720	12	BI147444	BI147444 602914008
33	348.6	62.8	915	10	BF166139	BF166139 601776586
34	344.6	62.1	652	13	BU352506	BU352506 603527982
35	344.2	62.0	650	9	AI858615	AI858615 wl40f01.x
36	342.6	61.7	639	12	BM998137	BM998137 UI-H-DT1-
37	341.4	61.5	614	13	BU459935	BU459935 603367327
38	341	61.4	634	13	BQ772348	BQ772348 UI-H-EZ1-
39	340.4	61.3	587	9	AA288198	AA288198 vb15b10.r
40	340	61.3	660	10	AW192502	AW192502 x145e10.x
41	339.2	61.1	690	12	BI219399	BI219399 602936756
42	336.2	60.6	795	14	CB601615	CB601615 AGENCOURT
43	335.2	60.4	634	9	AI970297	AI970297 wr09c02.x
44	334.8	60.3	706	13	BU439577	BU439577 603208252
45	333.8	60.1	745	10	AW243446	AW243446 xm95d11.x

ALIGNMENTS

```
RESULT 1
BF074459
LOCUS       BF074459                 551 bp    mRNA    linear    EST 25-APR-2001
DEFINITION  221883 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF074459
VERSION     BF074459.1  GI:10867970
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
ORGANISM    Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 551)
AUTHORS     Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
            Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
            Chitko-McKown,C.G., Pertea,G., Holt,I., Karamycheva,S., Liang,F.,
            Quackenbush,J. and Keele,J.W.
TITLE       Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
JOURNAL     Genome Res. 11 (4), 626-630 (2001)
MEDLINE     21180013
PUBMED      11282978
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
```

Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR PRimers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCCAGTCACGACG
Plate: 81 row: F column: 9
Seq primer: ATTTAGGTGACACTATAG.
FEATURES   Location/Qualifiers
     source   1..551
                   /organism="Bos taurus"
                   /mol_type="mRNA"
                   /db_xref="taxon:9913"
                   /tissue_type="pooled"
                   /lab_host="DH10B"
                   /clone_lib="MARC 2BOV"
                   /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
                   Library made from pooled tissue from testis, thymus,
                   semitendonosus muscle, longissimus muscle, pancreas,
                   adrenal, and endometrium."

ORIGIN
Query Match   79.2%; Score 439.8; DB 10; Length 551;
Best Local Similarity 89.3%; Pred. No. 3.8e-53;
Matches   474; Conservative   0; Mismatches   57; Indels   0; Gaps   0;

```
Qy   25 GTGGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCGGGGCGGCATGCGGAGGCATCCGGGGGA 84
           |||||||||||||||| |||||||||  |||| |||||||| ||||||  |||||||| |
Db    1 GTGGCTGCACCTGGTGGCGCTCAACAGCCCCGCAGTCGGGGCGGCCGCTCAAACAGCCCCGGGGC 60

Qy   85 GCGGGACTTCCAGTGCTTCAGCAGCGCGGCGCCGCGGGGCTGCTGGCCGCCGGCGACCTTCCGGGCC 144
           |||||||||||||||||||||||||  |||||||  |  |||||||
Db   61 GCCGACTTTCAGTGCTTCCAGCAGCGGGGCGGCGCTGGCGCGCGCGGCGCCGCCGCCGCGCCGCACCTTCCGCGCG 120

Qy  145 TTCCTGCGTCGTCGGCCTGCGGCTGCAGCAGGACCTCTACAGCATCGTGCGGCCGCCGACCGCACCGGG 204
           ||||||||| |||||||  ||||  ||||||||| || |||||||||||||||||||||||||||||  |
Db  121 TTCCTGCGTCCTGCGGGTTGCAGGACCATCGTACAGCATCGTGCGGCCGCCGACCGCACCGTGCCACC 180

Qy  205 GTGCCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTCCCCAGCTGGGAGGCCTTATTCTCG 264
           ||||||||| |||||||||||||||||||||||| ||| |||||||||||||||| ||||
Db  181 CTGCCCGGTCGTCAACCTCAGGGACGAGGTGCTGTTTCCTGGAGGCCTTGTTCTCA 240

Qy  265 GGCTCCGAGGGCCAGCTGAAGCCCCGGGGCGCCCGGGCCCGGGCGCTGGCGACCTGGGCGCGGGC 324
           |||||||||||||||||||||||||||  || |||||||
Db  241 GGCTCCGAGGGCCAGCTGAAGCCCGGGCCTGGCACGGCTCCGACGGGCTCCAGCAGGGGCG 300

Qy  325 CTGCAGCACCCCGCCTGGCCCCGGAAGAGAGCGTGTGGCACGGCTCCGAGGGCGGCGC 384
           || |||||  || |||||||||||||||||||||||||||||||||| |||||||||
Db  301 CTTCAGCATCCCACCTGGCCCCGGAAGAGAGCGTGTGGCACGGCTCCAGCCGGGGCGC 360

Qy  385 CGCCTGACCAGCAGCTACTGCGGACGTGGCGGACGTGCTGGCGGGCCCCGGCGGCCGCGGGCCACGGGCCAG 444
           || |||||| ||||||||||||||| ||| |||||| | |||  ||||||||||||  ||||||
Db  361 CGGCTGACCGGAGCTACTGCGGGACGTGGCGGACGTGGGCGGGCTGCTGGACAGCAGCAGGGGGCGGGGCCGCCACAGGCCAG 420

Qy  445 GCGTCGTCGCGTCGGCGGGCAGGGCGCTGCTGGAGCGGGCAGGAGGCCGCGGAGCTGCCGCCACCTGGCG 504
           || || ||||| ||||||| ||| ||||||||||||||||||   ||||||||  |||| ||||  |||
Db  421 GCCTCCTCGCGCTGCTGGCGGGGCCGGCGTTGCTGGAGCAGCAGAAAGCCGCGGGGGTGCCACAACGCC 480

Qy  505 TTCGTGGTGCTCTGCATCGAGAACAGCGTCATGACCTTCTCCTTCTCCAAGTAG 555
           |||  |||  ||||||||||||||||||  |||||||| |||| ||||||||||
Db  481 TTCATTGTCCCTGCATCGAGAACAGCTTCATGACCTCCTCCTCCAAGTAG 531
```

SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 707)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9704 row: d column: 16
            High quality sequence stop: 688.
FEATURES   Location/Qualifiers
     source   1..707
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
                   /db_xref="taxon:9606"
                   /clone="IMAGE:3902175"
                   /tissue_type="epithelioid carcinoma"
                   /lab_host="DH10B (phage-resistant)"
                   /clone_lib="NIH_MGC_70"
                   /note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI;
                   Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                   Average insert size 1.1 kb. Library constructed by Life
                   Technologies."

ORIGIN
Query Match   75.3%; Score 418; DB 10; Length 707;
Best Local Similarity 85.3%; Pred. No. 4.9e-50;
Matches   466; Conservative   0; Mismatches   80; Indels   0; Gaps   0;

```
Qy    1 CACACCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG 60
          ||| ||||| |||||||||| ||  || ||||||||| |||| ||||||||||  ||
Db   11 CACAGCCACCGCGACTTCCAGCCGCCGGTGGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA 70

Qy   61 GGCGGCATGCGAGGCATCCGGGAGGAGCGGACTTCCAGTGCTTCCAGCAGCGCGCGCCGCG 120
          |||||||||||  ||||||||  | |  |||||||||||||||||||||||  |||  || |
Db   71 GGCGGCATGCGGGGGCATCCGGCGGCCGACTTCCAGTGCTTCCAGCAGCGGGCGGCCGCCGTG 130

Qy  121 GGGCTGCCCGGCGCACCTTCCGTCGTCGCCGCCTTCCTGTCGTCGGCCTCAGGAGGACCATC 180
          |||||| ||||||||||||||||||||  |||||  | |||||||||||||||||||||||||
Db  131 GGGCTGCGCGGCGCACCTTCCGTCGTCCTGCGTCGGCCTCAGGAGGACCATC 190

Qy  181 GTGCGGCCGCCGACCGCACCGGGTGCCCGTGTCAACCTCAGGGACGAGGTGCTCTTC 240
          |||||||| || |||||||||||||||| || ||||||||| |||||| ||||| ||
Db  191 GTGCGGCCGTGCCGACCGCACCGGGTGCCCATCGTCAACCTCAAGGACGAGCTGCTCGTTT 250

Qy  241 CCCAGCTGGAGGCCTTATTCTCGGGCTCCGAGGGCCCAGCTGAGCCCCGGGCCCATC 300
          |||||||||  || ||| ||||| |||||||||| |  || ||  ||||||| || ||
Db  251 CCCAGCTGGGAGGCTCTGTTCCAGGCTCTGAGGGTCCGAGGGCTCGAGCCCGGGGCACGCATC 310

Qy  301 TTCTCTTTGACGGCGAGAGATGTCCTGCAGCACCCGCCCGGCCCCGGAAGAGAGCGTGTGG 360
          ||||| ||||||||||  || ||| ||||||  ||| || |||||  | ||| ||||||||
Db  311 TTCTCCTTTGACGGCGAAGGACGACGTCCTGAGGCACCCACCTGGCCCCAGTACTGCGAGACGTGTGTGG 370

Qy  361 CACGGCTTCCGACACCCCAGCGGCGCGGCGCAGCGTGACGGAGCTACTGCGGGGGGGACG 420
          || |||||||| ||||||| |||| ||| | |||| ||| |||| || | | |||||||
Db  371 CATGGCTTCCGGACACCCCAAGCGGCCACGGGCCGGCGCAGGCTGACCGGGAGCGGTGGGCGGGACG 430

Qy  421 GAGGCCCCGGCGGCCACGGCGCCGGCGTCGTGCGGCGTGCTGGCGGGCGTGCTGCTGGAGCAG 480
          ||||| |||| || ||||| | |   ||  |  |  |||||||  ||||||||||| |||||||
Db  431 GAGGCTCCCCGGCCACGGGCCCAGGCCGCTCCTGGCTGCTGGGGGCGGGGCGGGGCAGGGCCTCCTGGGGCAG 490

Qy  481 GAGGGCCGGAGCTGCCGCCACCTGGCGCATCGAGAACAGCGTCATGACC 540
          ||    ||| |||||| |||| ||||| |||||||||||| ||||||| |
Db  491 AGTGCCGGCGAGCTGCCATCACGGCCTACATCGTGCTTCTGCATTGGAACAGCGTTCATGACT 550
```

```
Qy   541 TCCTTC 546
         ||||||
Db   551 GCCTCC 556


RESULT 3
BE906253
LOCUS       BE906253                 757 bp    mRNA    linear    EST 20-OCT-2000
DEFINITION  601502237F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3904208 5',
            mRNA sequence.
ACCESSION   BE906253
VERSION     BE906253.1  GI:10399595
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 757)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9709 row: i  column: 09
            High quality sequence stop: 757.
FEATURES             Location/Qualifiers
     source          1..757
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3904208"
                     /tissue_type="epithelioid carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC 70"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.1 kb.  Library constructed by Life
                     Technologies."

ORIGIN
Query Match         75.3%;  Score 418;  DB 10;  Length 757;
Best Local Similarity   85.3%;  Pred. No. 5e-50;
Matches  466;  Conservative   0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy     1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGTGCCCTGAACAGCCCGCAGCCG   60
         |||||| ||||||||||||||||||||||||| ||  ||||||||||||||||||||||
Db    36 CACAGCCACCGGACTTCCAGCCGTGCTGCACCTGTTCCACACAGCCCCCTGTCA   95

Qy    61 GGCGGCATGCGGAGGCATCCGGGAGGCGGAGCGGCTTCCAGTGCTTCCAGCAGGGGCGCGGG  120
         ||||||||||||||||||||||||||||| || ||||||||||||||||||||||| ||||
Db    96 GGCGGCATGCGGGGGGCATCCGGCGGGGCCACTTCCAGCCGTTCCAGCAGGGGCGGGGCCGGTG  155

Qy   121 GGGCTGGCCGGCCGCCACCTTCCGGCCTTCGCGTCGTCCTGGCCTGCCGGGCTGCAGGGACCTTCACAGCATC  180
         ||||||||||||||| | | ||||||||| |||||||||||||||||||||||| ||||
Db   156 GGGCTGGCGCGGGCACCTTCCGGCCTTCCGCGCCTTCGTCCTGTCCTGCCGGGCCTGCAGGACCTGTACAGCATC  215

Qy   181 GTGGGGCCCGGCCGCACCGGACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC  240
         ||||||||||| |||||||||||| |||||||||||||||||||| ||||||||||| ||||
Db   216 GTGCGGCCTGCGCACCGGCAGCCGCGCAGCGCCGTGCCCATCGTCAACCTCAACGGACGAGCTGCTGTGTT  275

Qy   241 CCCAGCTCGGGAGGCCTTATTCTCGGGCTCCGAGAGGCCAGCTGAAGCCCGGGCCCGCATC  300
         |||||||| ||||||||||||| |||| |||||||||||||||||||||||| |||||||
Db   276 CCCAGCTCGGAGGGCTCTGTTCTTCAGGTCTGAGGGTCCGGGCTCGAAGCCCGGGGGCACGCATC  335

Qy   301 TTCTCTTTGACGGCAGAGATGTTCCTGAGCACCCCGCCTGGCCCCCGGAAGAGGCGTGTGG  360
```

TITLE   Direct Submission
JOURNAL   Submitted (08-DEC-2003) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
          USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Dr. James R. Lupski
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E Consortium (LLNL)
          DNA Sequencing by: Genome Sequence Centre,
          BC Cancer Agency, Vancouver, BC, Canada
          info@bcgsc.bc.ca
          Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
          Susanna Chan, Redman Chiu, Chris Fjell, Erin Garland, Ran Guin,
          Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
          Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
          Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
          Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
          Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
          George Yang, Scott Zuyderduyn, Marco Marra.

          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
          Series: IRAK Plate: 141 Row: c Column: 6
          This clone was selected for full length sequencing because it
          passed the following selection criteria: matched mRNA gi: 18765747
          This clone has the following problem: frame shifted.
FEATURES   Location/Qualifiers
     source   1..4230
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6181818"
                     /tissue_type="Peripheral Nervous System, dorsal root
                     ganglion"
                     /clone_lib="Lupski_dorsal_root_ganglion"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
ORIGIN

Query Match   75.3%;   Score 418;   DB 11;   Length 4230;
Best Local Similarity 85.3%;   Pred. No. 6.7e-50;
Matches 466; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy   1  CACACCCACCAGGACTTCCAGTGGTGCTGCACCTGGTGCCCTGAACAGCCCGCAGCCG   60
        ||||| ||| | ||||||||||  | ||||||||||| |||| |||||||||| |
Db 3495 CACACGCACCCGCGACTTCCAGCCCGTGCTGCACCTGGTTGCGCTCAACAGCCCCCTGTCA 3554

Qy  61  GGGGCATGCGAGGAGCATCCGGGGAGGCGACTTCCAGTGCTTCCAGCAGGCGCGCCGCG  120
        ||||||||||  ||| ||||| ||| ||| ||||||||||||||||||||||||  ||| |
Db 3555 GGGGGCATGCGGGGCATCCGCGGGCCGGGCACTTCCAGTGCTTCCAGCAGGCGCGGCCGTG 3614

Qy 121  GGGGTGGCGCGGCACCTTCCGGTGCGCAGGCTGCGCAGGTGCGCAGGACCTCTACAGCATC 180
        |||||||||| ||||||||| | | || | | | ||| || |||||||| ||||||||||
Db 3615 GGGGTGGCGGGGCACCTTCCGCGGCTTCCGTGCTCCGGCCGTGCGAGGCCTGTACAGCATC 3674

Qy 181  GTGGCGCCGGCGACACCGCACCGGGGTGCCCGTCGTCAACCTCAGGGAGGAGGTGCTCTTC 240
        |||||||| | || || | ||||| |||||| ||||||||||| |||| ||| ||| |
Db 3675 GTGGCGCGGTGCCGACCGGCAGCCGTGCCCATCGTCAACCTCAAGGACCAGCTGCTGTT   3734

Qy 241  CCCAGCTGGGAGGCCTTATTCTCGGGTCCGAGGGGTGAAGCCCGGGCCGCATC        300
        |||||||||||||| | |  |||| |  | |||||||  |||||||| | ||||
Db 3735 CCCAGCTGGGAGGCTCTGTTCTCTCAGGCTCTGAGGGGTCCGGGAAGCCCGGGCACGCATC 3794

Qy 301  TTCTCTTCGACGGCAGAGATGCCTGCAGCACCCCGGCTCGCCCCGGGAAGAGGGTGTGG   360
        ||||| |||||||||  || |||  |||| | |||| |  |  |||||||| | ||||||
Db 3795 TTCTCCTTTGACGGCAAGGACGTCCTGAGGGCCCCGGAAGAGGCTGTGG            3854

Qy 361  CACGGGTCCGACCCCAGCCGGCGGCGCCGCCTGACCGACCAGCTACTGCCGACACCGTGGCGACG 420
        || ||| |||||||||| ||| ||| | |  ||| ||  | || ||| |  || |   ||| ||||
Db 3855 CATGGCTCGGACCCCAACGGCGCAGGCGGCACTGTGAGACGTGGGGGACG          3914
```

Db 352 GGGGCATGCGGGGCATCCGCGGGCCGGGCACTTCCAGTGCTTCCAGCAGGCGCGGCCGTG 411
Qy 121 GGGCTGGCGCCGGCACCTTCCGGTGCGCACGGCCGTTCGCGGCCGTGCGCAGGACCTCTACAGCATC 180
Db 412 GGGCTGCGGGGCACCTTCCGCGCGTTCCCGCCGCTGCGCAGGACCTGTACAGCATC 471
Qy 181 GTGGCGCCGCCGACCGCCACCGGGGTGCCGTCGTCAACCTCAGGAGGACGAGGTGCTCTTC 240
Db 472 GTGGCGCGGTGCCGACCGGCGCAGCGCGTGCCCATCGTCAACCTCAAGGACGAGCTGCTGTTT 531
Qy 241 CCCAGCTGGGAGCGCCTTATTCTCGGGTCCGAGGGCGTCCGAGGGCTGAAGCCCGGGCCCGCATC 300
Db 532 CCCAGCTGGGAGGCTCTGTTCTCTCAGGCTCTGAGGGTCCGGGAAGCCCGGGCACGCATC 591
Qy 301 TTCTCTTCGACGGCAGAGATGCCTGCAGCACCCCGGCTCGCCCCGGGAAGAGGGTGTGG 360
Db 592 TTCTCCTTTGACGGCAAGGACGTCCTGAGGGCACCCCACCTGGCCCCAGAAGAGGCGGTGTGG 651
Qy 361 CACGGGTCCGACCCCAGCCGGCGGCGCCGCCTGACCGACCAGCTACTGCCGACACCGTGGCGACG 420
Db 652 CATGGCTCGGACCCCAACGGCGCAGGCGGCACTGTGAGACGTGAGCGGCGGACG 711
Qy 421 GAGGGCCCCGGCGGCGCACCGGGCAGGCGTGCTCGTCGGCCGGCAGGCTCGTCTGGAGCAG 480
Db 712 GAGGGCCCCTCGGCCCACGAGGCCAGGCCTCCTCGTCGTGCTGGGGGCGGCAGGCTCCTGGGGCAG 771
Qy 481 GAGGGCGGCGAGCTGCGCCCACGCCTTCGTGGTGCTCTGCATCGAGAACAGCGGTCATGACC 540
Db 772 AGTGCCGCGAGCTGCCATCACGCGGCCCTACATCGTGCTCTGCATTGAGAACAGCTTCATGACT 831
Qy 541 TCCTTC 546
        ||||||
Db 832 GCCTCC 837

RESULT 5
BC063833
LOCUS   BC063833   4230 bp   mRNA   linear   HTC 09-DEC-2003
DEFINITION Homo sapiens cDNA clone IMAGE:6181818, containing frame-shift
            errors.
ACCESSION   BC063833
VERSION   BC063833.1 GI:3645297
KEYWORDS   HTC.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 4230)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2 (bases 1 to 4230)
  AUTHORS   Strausberg,R.

Mon Mar 29 17:32:22 2004

```
       181 GGGGCGGGGAGGCCTTGTTCTCGGGCTCTGAGGGCTTCTGGGGGCCAGCTGAGGGCCCGGCGGCCGGGCCGCTGAAGCCCGGGGCCCTGAAGCCTCAGCATCT 240
                                                                                                                
       302 TCTCTTTCGACGGCCAGAGATGTCCTCGACACCCGGCTCCTGCCCCTGGCCCCCGGAGAGCGTGTGGC 361
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       241 TCTCTTTCGACGGCCAGAGACGTTCCTTCAGCACCCTGCCTCTGGCCCCAGAGAAGAGCGTGTGGGC 300
```

```
       362 ACGGCTCCGACCCCAGCCCGCCGCCGGGCGGCGCAGGGCCAGGGGCCAGGTACTGCGAGACGCGG 421
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       301 ACGGCTCAGACCCCAGCCCGGCGCCGGCGCCGGCCACTGCTGACCGAGAGCTACTGTGAGACGTG 360
```

```
       422 AGGCCCCGGCGGCCACCGGGCAGGGCAGGGCCGTCGTGCTGCTGGCCGGGCAGGCTGCTGGAGCAGG 481
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       361 AGGCCAGGACGGGCCACGGGCCAGGCCTCCTGCTGGCCGGCGGCCGGCTGCTGCTGGAGCAAA 420
```

```
       482 AGGGCGGAGCTGCCGCCACAGCCGCCTTCGTGCTGTGGTGCTTCTGCATCGAGACAGACAGCGTCATCGACCT 541
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       421 AGGGCGGAGCTGCCGCTGCCACCACGCCTTCATTGTTCTGTGCATCATCGAGACAGCAGCTTCATGACCT 480
```

```
       542 CCTTCTCCAAGTAG 555
           ||||||||||||||
       481 CCGGCTCCAAGTAG 494
```

RESULT 7
BU859398/c
LOCUS   BU859398   944 bp   mRNA   linear   EST 16-OCT-2002
DEFINITION   AGENCOURT_10434549 NIH_MGC_109 Homo sapiens cDNA clone
             IMAGE:6650260 5', mRNA sequence.
ACCESSION   BU859398
VERSION   BU859398.1   GI:24044390
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 944)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM2895 row: d column: 04
             High quality sequence stop: 572.
FEATURES   Location/Qualifiers
     source   1..944
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6650260"
                     /tissue_type="teratocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC 109"
                     /note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
                     XhoI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN

Query Match   71.0%;   Score 394.2; DB 13; Length 944;
Best Local Similarity 83.7%; Pred. No. 1.2e-46;
Matches 458; Conservative 0; Mismatches 88; Indels 1; Gaps 1;
```

---

```
       421 GAGGCCCCGGCGGCCACCGGGCAGGGCGGCTCGTCGCTGCTGGCGGGCAGGCTGCTGGAGCAG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      3915 GAGGCTCCTCGGCCCACGGGCCAGGCCGGCCAGGGCCCAGGGCCAGGCAG 3974
```

```
       481 GAGGCCGCGAGCTGCCGCCACGCCTTCGTGCTTCTGTGGTGCTTCTGCATCGAGACACAGCGTCATGACC 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      3975 AGTGCCGCGGAGCTGCCATCACGCCTACACGTGTCTCTGCATTGAGAACAGCAGCTTCATGACT 4034
```

```
       541 TCCTTC 546
           ||||||
      4035 GCCTCC 4040
```

RESULT 6
CF789984
LOCUS   CF789984   504 bp   mRNA   linear   EST 21-OCT-2003
DEFINITION   867322 MARC 4PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   CF789984
VERSION   CF789984.1   GI:37794545
KEYWORDS   EST.
SOURCE   Sus scrofa (pig)
  ORGANISM   Sus scrofa
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1 (bases 1 to 504)
  AUTHORS   Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,
             Nonneman,D.J., Wray,J.E. and Keele,J.W.
  TITLE   Porcine EST collection using a normalized library constructed from
             embryos representing early developmental stages
  JOURNAL   Unpublished (2003)
COMMENT   Contact: Smith TPL
             USDA, ARS, US Meat Animal Research Center
             PO Box 166, Clay Center, NE 68933-0166, USA
             Tel: 402 762 4366
             Fax: 402 762 4390
             Email: smith@email.marc.usda.gov
             Single pass sequencing. Bases called with phred v0.020425.c and
             trimmed with the aid of the trim_alt option. Vector identified with
             cross_match v0.990329.
             Plate: TMW8006 row: B column: 3
             Seq primer: GTAATACGACTCACTATAGGG.
FEATURES   Location/Qualifiers
     source   1..504
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 4PIG"
                     /note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
                     Library made with combined RNA from day-10, day-13,
                     day-15, day-25, and day-30 whole embryos."
ORIGIN

Query Match   74.0%;   Score 410.8; DB 14; Length 504;
Best Local Similarity 89.5%; Pred. No. 4.9e-49;
Matches 442; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
```

```
        62 GCGGCATGCGAGGAGGCATCCGGGGAGCGGACTTCCAGTGCTTCCAGCAGGCGGCGGCCGCGCGG 121
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         1 GCGGCCTGGCGCGGCCATCCGCGGCGCGCGACTTCCAGTGCTTCCAGCAGGCGCCGCCGCCTGG 60
```

```
       122 GGCTGGCCGGGCACCTTCCCGGGCTTCCCGGTGTCGTCGGGCGGCCTGACGGCCTGGCGGCCATCG 181
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        61 GGCTGGCGGGGCACCTTCCCGGGCCCTTCCTGTCGTCGTCGCGCGGCCTGACGGCCTGGCGGCCATCG 120
```

```
       182 TGCGCGGCCGCGACCACGGGGTGCCCGTGCCGTCCAACCTCAGGGACGAGGTTGCTCTTCC 241
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       121 TGCGACGGCCGCGACCGCGCCCGTGCCCATCGTTAACCTCAGGGACGAGGTGCTGTTCG 180
```

```
       242 CCAGCTGGGAGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGGCCGCATCT 301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy      1 CACACCCACCAGGACTTCCAGCTGGT-GCTGACCTGGTGGCCCTGAAACAGCCCGCAGCC 59
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db    712 CAGGTTCGACCGAGGCTTCCAGCCGGTGGGTCCACATGGTTGCGCTCAGCAGCCCCCTGTC 653

Qy     60 GGGGCCGCATGCGAGGCATCCGGGAGGAGGAGCTTCCAGTGCTTCCAGCGAGCGCGCCGC 119
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    652 AGGGGCCGCATGCGGGGCATCCGCGGGCCCGACTTCCAGTGCTTCCAGCGCGCGCGCCGT 593

Qy    120 GGGGGCTGGCCGGCACCTTCCGGGACCTTCCGTGCGCGGGTGCAGGACCTTCACAGCAT 179
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    592 GGGGGCTGGCCGGGCCACCTTCCGCGCCTTCCTGTCCTCCGCGCTCAGGACCCTGTACAGCAT 533

Qy    180 CGTGGCGGCGCCGCCGCCAGCGCGCGGGTGCCGTCGTCAACCTCAGGACGAGGGTGCTCTT 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    532 CGTGGCGCCCGTGCCGACCGCGGCAGCCGTGCCCATCGTCAACCTCAAGGACGACCAGCTGCTGTT 473

Qy    240 CCCCAGCTCGGGAGGCCTTATTCCGGGCTCCGAGGCCCAGTGAAGCCCGGGGCCCGCAT 299
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    472 TCCCCAGCTCGGGAGGCTCGTTCTTCCAGGGGTCTCGGTCAGGGTCCGCTGAAGCGCACGCAT 413

Qy    300 CTTCTCTTTCGACGGCAGAGATGTCCTGCAGCACCCCGGCCCGGGCGAAGAGCGTGTG 359
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    352 CTTCTCCTTTGACGGCAAGGACGTGCTGAGGACCAGCACCCCACCTGGCCCCACCTGGCTGTG 353

Qy    360 GCACGGCTCCGACCCCAGCGGCGGCCGCCTGACCGACAGCTACTGCGAGCGTGCGGAC 419
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    352 GCATGGCTCGGACCCCAACGGGCGCCAGCGGCGCAGGCTACTGTGACGGTGGCCGGAC 293

Qy    420 GGAGGCCCCGGCGGCACCGGAGCAGGCGTCGTCGTGCTGCTGGGGGGGGGCAGGGCA 479
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    292 GGAGGCTCCCTGGCCACGGGCCAGGCCTCCTCGGCGTCCTGTGGGGTGCGGGTCCTGGGGGCA 233

Qy    480 GGAGGCGCCGAGCTGCCGCGCACCGCTTCGTGGTGCTCTGCACAGCGTCATGAC 539
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    232 GAGTGCCGCGAGCTGCCGCAGCTGCCATCACGCTCACATCACCTGCTCTGCAATGAGCAACAGCTTCATGAC 173

Qy    540 CTCCTTC 546
          |||||||
Db    172 TGCCTCC 166
```

```
RESULT 8
CB444165/c
LOCUS        CB444165            657 bp    mRNA      linear   EST 25-MAR-2003
DEFINITION   695295 MARC 6BOV Bos taurus cDNA 3', mRNA sequence.
ACCESSION    CB444165
VERSION      CB444165.1  GI:29233914
KEYWORDS     EST.
SOURCE       Bos taurus (cow)
  ORGANISM   Bos taurus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
             Bovidae; Bovinae; Bos.
REFERENCE    1  (bases 1 to 657)
  AUTHORS    Smith,T.P.L., Roberts,A.J., Echternkamp,S.E., Chitko-McKown,C.G.,
             Wray,J.E. and Keele,J.W.
  TITLE      A second set of bovine ESTs from pooled-tissue normalized libraries
  JOURNAL    Unpublished (2003)
COMMENT      Contact: Smith TPL
             USDA, ARS, US Meat Animal Research Center
             PO Box 166, Clay Center, NE 68933-0166, USA
             Tel: 402 762 4356
             Fax: 402 762 4390
             Email: smith@email.marc.usda.gov
             Single pass sequencing. Bases called with phred v0.020425.c and
             trimmed with the aid of the trim_alt option. Vector identified with
             cross match v0.990322.
             Plate: FQ78051 row: I  column: 11
             Seq primer: TAGAAGGACCAGTCGAGG.
FEATURES             Location/Qualifiers
     source          1..657
                     /organism="Bos taurus"
                     /mol_type="mRNA"
```

```
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 6BOV"
                     /note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
                     Library made with RNA pooled from multiple tissues
                     including liver, lung, hypothalamus, pituitary, and
                     placenta/endometrium."

ORIGIN

Query Match              70.6%;  Score 391.8;  DB 14;  Length 657;
Best Local Similarity    88.9%;  Pred. No. 2.5e-46;
Matches  423;  Conservative   0;  Mismatches  53;  Indels   0;  Gaps   0;
```

```
Qy     80 GGGGAGCGGACTTCCAGTGCTTCCAGCAGGCGGCGGCGGGGCTGGCCGGGGCCACCTTCC 139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    657 GCGGCGCGGACTTCAGTGCTTCCAGCGCGCGCGCCGTGGGGGCTGGCCGGGCCACCATTCC 598

Qy    140 GGGGCCTTCCTGTCGTCGGCTGCGGCGACCTTCACAGCATCGTGGCGGCCGCGCGACCGCA 199
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    597 GCGCGGTTCCTGTCCTCGGCGTTGCAGGACCTTGCAGGACCATCGTGCGCGCCGGCACGCCGTG 538

Qy    200 CCGGGGTGCCCGTCGTCAACCTCAGGGACGACGAGTGCTCTTCCCAGCTGGGAGGCCTTAT 259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    537 CCACCCGCCCGTCGTCAACCTCAGGGACGAGGGTGCTGTTTCCTAGCTGGGAGGCCTTGT 478

Qy    260 TCTCGGGCTCCGAGGGCCAGCTGAAGCCCGGGGCCGCCCGCCATCTTCTCTTTCGACGGCAGAG 319
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    477 TCTCAGGCTCCGAGGGGCAAGCTGAAGCCCGGGCGCCCCCATCTTCTCCTTCGACGGCAGAG 418

Qy    320 ATGTCCTGCAGCACCCCGGCCCGGGCCCCGGAAGAGCGTGTGGCACGGCTCCGACCCCAGCG 379
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    417 ATGTCCTTCAGCATCCCACCTGGCCCCACCTGGCCGTGGGACAGGCTCGGACCCCAGCG 358

Qy    380 GCGCGCGCCTGACCGACAGCTACTGCGAGACGTGGCGACGGAGGCCCCGGGCGGCCACCG 439
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    357 GCGCGCGGCTGACGGAGAGCTACTGCGAGACGTGGCGAGACGTGGCGACGCAGGGGGCGCCACCGG 298

Qy    440 GGCAGGCGGTCGTCGTGCTGCTGGCGGGGCCGCTGCTGAGGCAGGGCAGCTGCCGCGCC 499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    297 GCCAGGCCTCCTCGGCGTCGTCGGGCGGGCCGCTGCTGTGGGAGCAGAGAAAGCCGGGGTGTGCCACA 238

Qy    500 AGCCTTCGGTGGTGCTTCGCATGAGAACAGCGTCATGACCTTCCTTCCTTCCCAAGTAG 555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    237 ACGCCTTCATTGTCCTNTGCATCGAGAACAGCTTCATGACCTCTCCTCCCAAGTAG 182
```

```
RESULT 9
BU615520/c
LOCUS        BU615520            703 bp    mRNA      linear   EST 23-SEP-2002
DEFINITION   UI-H-FG0-bcp-p-10-0-UI.s1 NCI_CGAP_EN1_2 Homo sapiens cDNA clone
             UI-H-FG0-bcp-p-10-0-UI 3', mRNA sequence.
ACCESSION    BU615520
VERSION      BU615520.1  GI:23281735
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 703)
  AUTHORS    Strausberg,R.
  TITLE      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
             National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: James Marrin
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be obtained
             from Dr. M. Bento Soares, bento-soares@uiowa.edu
             The following repetitive elements were found in this cDNA
```

```
LOCUS       BQ672290                 929 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8354946 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6275261
            5', mRNA sequence.
ACCESSION   BQ672290
VERSION     BQ672290.1  GI:21783124
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 929)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2458 row: c  column: 06
            High quality sequence stop: 588.
FEATURES             Location/Qualifiers
     source          1..929
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6275261"
                     /tissue_type="epidermoid carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_102"
                     /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of California, Berkeley) using ZAP-cDNA
                     synthesis kit (Stratagene) and Superscript II RT (Life
                     Technologies). Note: this is a NIH_MGC Library."
ORIGIN
```

```
Query Match            69.2%;     Score 384.2;   DB 13;   Length 929;
Best Local Similarity  82.5%;     Pred. No. 3.2e-45;
Matches  453;  Conservative    0;  Mismatches  93;  Indels   3;  Gaps    1;

Qy        1 CACACCCACCAGAGACTTCCAGCTGGTGCTGCCACCTGGTGGCCCTGAACAGCCCGCAGCCG   60
            ||| ||||  ||||| |||||||||| |||||||||||| |||||||||||| ||||||
Db      227 CACAGCCACCCGGACTTCCAGCCGTGCTTCCACCTGGTTGCGCTCAACAGCCCCCTGTCA  286

Qy       61 GGGGCATGCGGAGGCATCCGGGGGACCTTCCGGGGAGCGGGGGACTTCCAGTGCTTCCAGTGCTTCCAGTGCGCCGCGCG  120
            |||||||||||||||  |||||||||||||||||||| || || |||||||||||||||||||
Db      287 GGGGCATGCGGGGGCATCCGGGGGCCCGGGGCGCCGGGGGGCGCGGGGGCCGTG  346

Qy      121 GGGCTGCCGCGCCACCTTCCGGGGCCTTCGTCGTGCTCCCGTCTCCGTGCCCCGTGTCGTCCCGTCGTCAAC  180
            |||||||||||||||||||||||||| |||| |||||||||| ||||||||||||||||
Db      347 GGGCTGGCGGGCACCTTCCGGGGCCTTCCGTCCTCCGTGTCCTGCCAGGACCTGTACAGCATC  406

Qy      181 GTGCGGCCGCGGCCGACCGGCACCGGGGGTGCCCGGTGCTCAACCTCAGGCTCTGAGGGTCCG  240
            |||||||||||||||||||||||| |||||||||||||| |||||||||||||||||||||||||
Db      407 GTGCGGCCGTGCCGACCGCGCAGCCGTGCCATCGTGCCAACCTCAAGGACCAGAGCTGCTTCTT  466

Qy      241 CCCAGCTGGAGGCCTTATTCTCGGGCTCCGAGGGCCATCTGGGCTCCAGGCTCTGAGGGCCATC  300
            |||||||||||||||  || |||||||||| |||||||||||||||||||
Db      467 CCCAGCTGGGAGGCTTCTGTTCCGAGCTCTGAGGGTCCCGCTGCTGGCCCGGCCACGCAGCATC  526

Qy      301 TTCTCTTTCGACGGACGAGAGAGATGTCCTGAGCAGCAGTGCCCGGAGAGGCGTGTGG  360
            |||| || ||||||||| |||||
Db                                                                   360
```

```
sequence: 1-44, >POLY_A#Simple_repeat (matched compliment)
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES             Location/Qualifiers
     source          1..703
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-FGO-bcp-p-10-0-UI"
                     /tissue_type="Enchondroma cell line"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_EN1_2"
                     /note="Organ: Bone; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site 1: EcoR I; Site_2: Not I;
                     NCI_CGAP_EN_2 is a cDNA library containing the following
                     tissue(s): Enchondroma cell line (2 cell lines). The
                     library was constructed according to Bonaldo, Lennon and
                     Soares, Genome Research, 6:791-806, 1996. First strand
                     cDNA synthesis was primed with an oligo-dT primer
                     containing a Not I site. Double stranded cDNA was ligated
                     to an EcoR I adaptor, digested with Not I, and cloned
                     directionally into pT7T3-Pac vector. The oligonucleotide
                     used to prime the synthesis of first-strand cDNA contains
                     a library tag sequence that is located between the Not I
                     site and the (dT)18 tail. The sequence tag for this
                     library is CCGGTCACTC. The cell lines was provided by Dr
                     James Martin from University of Iowa.
                     TAG_TISSUE=Inchondroma cell line (Mix of EN1 and EN2)
                     TAG_LIB=UI-H-FGO
                     TAG_SEQ=CGGTCACTC"
ORIGIN
```

```
Query Match            69.3%;     Score 384.4;   DB 13;   Length 703;
Best Local Similarity  84.8%;     Pred. No. 2.9e-45;
Matches  430;  Conservative    0;  Mismatches  77;  Indels   0;  Gaps    0;

Qy       40 GCCCTGAACAGCCCGCAGCCGGGGCATGCGGAGGCATCCGGGGGAGCGGACTTCCAGTGC   99
            ||| || ||||||||| ||||| |||||||||||||||| ||||||||| |||||||||||
Db      699 GCGCTCAACAGCCCCCTGTCAGGGGCATGCGGGGGCATCCGCGGGCGGACTTCCAGTGC  640

Qy      100 TTCCAGCAGGCGCCGCGCGGGGCTGCCGGCGGGGGCTGGCCGGCCACCTTCCGGGGTCGTCGCGG  159
            ||||||||||||  ||||||||| ||||||||||||| || |||||||||||||||||||||
Db      639 TTCCAGCAGGCGGCGCGGCGGGNCCGTGGGGCTGGCGGGCTGGCGGGCACCTTCCGGGGTCCTCGCGC  580

Qy      160 CTGCAGGACCTCTACAGACATCGTGCGGCCGCGGCCGACCGGCACCGGGGTGCCCGTCGTCAAC  219
            |||||||||||||||||||||||||||||||| |||||| || |||| ||||||| ||||||
Db      579 CTGCAGGACCTGTACAGCATCGTGCGGCCGTGCCGACCGCGCAGCCGTGCCATCGTCAAC  520

Qy      220 CTCAGGGACGAGGTGCTCTTCCCAGTGGGAGGCCTTATTCTCGGGCTCTGAGGGGTCCG  279
            ||||||| |||||||||||||||||||| ||||||||||  ||||||||||||| ||||
Db      519 CTCAAGGACGAGCTGCTGTTTCCCAGCTGGGAGGCTCTCGTTCTTCAGGCTCTGAGGGGTCCG  460

Qy      280 CTGAAGCCCGGGGCCCCGCCCATCTTCTTTGACGGCAGAGATGTCCTGCAGCACCCGCC  339
            |||||||||||||||  |||| ||||| || ||||||||||||||||| ||||| ||||
Db      459 CTGAAGCCCGGGGCAGCGCCATCTCTCTTTGACGGCAAGGACGTCCTGAGGCACCCCACC  400

Qy      340 TGGCCCCCGGAAGAGAGCGTGTGGCACGGGCCTCCGACGGGCGCCGCTGACCGACAGC  399
            |||||||||||||||||||| || |||||||||  ||||| |||||||||||||||||
Db      399 TGGCCCCAGAAGAGAGCGTGGCATGGGCCACGGGCGCCACCGGGCGGCCAGGCTGACCGAGAGC  340

Qy      400 TACTTGCGAGACGTGCCGACGACGGACGGAGGCCCACCGGCCACCGGGGTCGTCGCTG  459
            |||| |||||||||||||||||||| ||||||||||||| ||||||||| ||||||||
Db      339 TACTGTGAGACGTGCCGACGGGCGGACGGAGGCCCGACCGGCCAGGGCTCCCTGCGTCGGTG  280

Qy      460 GCGGGCAGGCTTGCTGGAGACAGGAGGGCCGGCGCGGAGCTGCGCCGCCACCGCCTTCGTGTG  519
            |||||| ||| || ||| |||||||||||| ||||||| ||||||| ||||||||||| |||||
Db      279 GGGGGCAGGCCTCTTGGGGCAGAGTGCGGGACGAGAGTGCCATCACGCCTACAGTCGCTCTGC  220

Qy      520 ATCGAGAACAGCGTCATGACCTCCTTC  546
            ||||||| || |||||||||| |||
Db      219 ATTGAGAACAGCTTCATGACCTGCCTCC  193
```

```
Db   527 TTCTCCTTAACGGCAAGAGACGTCCTGAGGCACCCACCTGGCCCCCAGAAGAGCGTGTGG 586
Qy   361 CACGGCTTCGGACCCCAGCGGCGGCGCGCGCCGCTTGACCGACAGCTACTGCGAGACGTGGCCGACG 420
         |||||||||||||||||| ||||||||||||| ||||||||||
Db   587 CATGGCTCGGACCCCAACGGCGGCGCCAGGCTGGCCGAGCAGCTACTGTGAGACGTGCCGGACG 646
Qy   421 GAGGGCCCGGCGGCCACCGGGGCCAGCGGCGTGCTGCGGCGGG---CTGCTGGAG 477
         ||||||||||||||||||||| ||| ||| |||||||||
Db   647 GAGGGCTCCCTCGGCCACGGGCCAGCCCTTCCTCCTCCCCTGCTGGGGGGGGCAGGCTCCCTGGGGC 706
Qy   478 CAGGAGGCCCGAGCTGGCCGCCCACGCCTTCGTGGTGCTTCTGCATCCGAGAACAGCCGTCATG 537
         |||||||||| ||||||||||||||||||||||| || |||||
Db   707 AGAGTGCCCGCGAGCTGGCCATCATCACGCCCTACATCGTGCTTCTGCATTGAGAACCGCGCTTCATG 766
Qy   538 ACCTCCTTC 546
         ||||||||
Db   767 ACTGCCTCC 775
```

RESULT 11
BM683067/c
LOCUS   UI-E-EO1-ajd-g-02-0-UI.s1   683 bp   mRNA   linear   EST 27-FEB-2002
DEFINITION UI-E-EO1-ajd-g-02-0-UI UI-E-EO1 Homo sapiens cDNA clone
            UI-E-EO1-ajd-g-02-0-UI 3', mRNA sequence.
ACCESSION   BM683067
VERSION   BM683067.1   GI:18992963
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 683)
AUTHORS   Bonaldo.M.F., Lennon,G. and Soares.M.B.
TITLE   Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   97044477
PUBMED   8889548
COMMENT   Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
            cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            The following repetitive elements were found in this cDNA
            sequence: 1-44, >POLY A#Simple_repeat (matched compliment)
            Seq primer: M13 Forward
            POLYA=Yes.

FEATURES   Location/Qualifiers
     source   1..683
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-E-EO1-ajd-g-02-0-UI"
                     /tissue_type="fetal eye"
                     /dev_stage="fetal"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-E-EO1"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site 1: EcoR I; Site 2: Not I;
                     UI-E-EO1 is a normalized cDNA library containing the
                     following tissue(s): fetal eye. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     CGCGATACC. This library was created for the program, Gene
                     Discovery in the Visual System, supported by National Eye
                     Institute (NEI).
                     TAG_TISSUE=human fetal eye
                     TAG_LIB=UI-E-EO1
                     TAG_SEQ=CGCGTATACC"

ORIGIN

Query Match   67.9%;   Score 377; DB 12;   Length 683;
Best Local Similarity   85.5%;   Pred. No. 3.2e-44;
Matches   419; Conservative 0; Mismatches 71; Indels 0; Gaps   0;

```
Qy    57 GCCGGGCGGCATGGAGGCATCCGGGAGGCGACTTCCAGTGCTTCCAGCCAGCGCGCGC 116
         |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   682 GTCAGGCGGCATGGAGGCATCCGGGGGGCCGACTTCCAGTGCTTCCAGCCAGCGCGCGGNC 623
Qy   117 CGCGGGGCTGGCCGGCCACCTTCCGGGGCCTTCCGTGCGGCTGCAGGGACCTCTACAG 176
         || ||||| |||||| |||||||||||||||||||| ||||||||||||||| |||||
Db   622 CGTGGGGCTGGCGGGCCACCTTCCGGGGCCTTCCGTGCTGCGGCTGCAGGGACCTGTACAG 563
Qy   177 CATCGTGCGCGCCGCCGACCGCACCGGGGTGCCCGTCGTCAACCTCACGGGAGGAGGTGCT 236
         ||||||||||| |||||||||| ||| ||||||||||| ||||||||||||||||| ||||
Db   562 CATCGTGCGCGTGCCGACCGCGCAGCCCGTGCCCATCGTCAACCTCACGGGAGGAGCTGCT 503
Qy   237 CTTCCCCAGCTGGGAGGCCTTATTCTCGGGCTCCGAGGGCCAGTGAAGCCCGGGGCCCG 296
         ||||||||||||||||||| | ||||| ||||| ||||| |||  |||||||||||||
Db   502 GTTTCCCAGCTGGGAGGCTCTGTTCTCAGGCTCTGAGGGTCCGCTGAAGCCCGGGGCCACG 443
Qy   297 CATCTTCTTTGACGGCAGAGATGTCCTGCAGCACCCCGGCTGGCCCGGAAGAGCGT 356
         |||||||| |||||||| || |||||||| ||||||| || |||| || ||||||||
Db   442 CATCTTCTCCTTTGACGGCAAGGACGTCCTGAGGCACCCCACCTGGCCCCCAGAAGAGCGT 383
Qy   357 GTGGCACGGCTTCCGACCCCAGCGGCGGCGCGCGCCTGACCGAGCACTGCGAGACGTGGCG 416
         |||||| ||||||| |||  |||| ||||| ||| ||| |||||||| ||| |||||||||
Db   382 GTGGCATGGCTTCGGACTTCGGACCCCAACGGCGGCCAGGCTGGACCGAGCAGAGCTACTGTGAGACGTGGCG 323
Qy   417 GACGAGGGCCCCGGCGGCCACCGGCGCAGGCGTGGTGTGCTTGGCGGGCAGGCGTGCTGGA 476
         |||| ||||| |||||||||  ||| |||  ||||||||  |||||||||||||| |||||
Db   322 GACGGAGGCTCCCTCGGGCCACGGGCCCAGCCTCCTCGCCTGCTGCTGGGGGGGGCAGGCGTCCTGGG 263
Qy   477 GCAGGAGGCGCCGGAGCTGGCCGCCGCCACGCCTTGGTGGTGCTCTGCATCGAGAACAGCGTCAT 536
         ||| || |||| ||||||||||  || || |||||| ||||||| ||||||| |||||||| |||
Db   262 GCAAGAGTGACGCGCCGAGCTGGCCATCACACGCCTACATCGTGCTTCTGCATTGAGAACAGCTTCAT 203
Qy   537 GACCCCTTC 546
         |||  ||||
Db   202 GACTGCTCC 193
```

RESULT 12
AW089583/c
LOCUS   AW089583   682 bp   mRNA   linear EST 15-OCT-1999
DEFINITION xd20f03.x1 NCI_CGAP Ov23 Homo sapiens cDNA clone IMAGE:2594333 3'
            similar to SW:CAIH_HUMAN P39060 COLLAGEN ALPHA 1(XVIII) CHAIN
            [CONTAINS: ENDOSTATIN] ;, mRNA sequence.
ACCESSION   AW089583
VERSION   AW089583.1   GI:6046927
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 682)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.

```
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: polyT not found
Seq primer: -40UP from Gibco
High quality sequence stop: 408.

FEATURES             Location/Qualifiers
     source          1..682
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2594333"
                     /tissue_type="tumor, 5 pooled (see description)"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Ov23"
                     /note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.35 kb. Tumor types include: mixed
                     Mullerian tumor, papillary serous, clear cell, spindle
                     cell.  All are primary tumors, metastasis positive. Life
                     Technologies catalog #: 11534-013"

ORIGIN

Query Match          67.9%;   Score 376.6; DB 9;  Length 682;
Best Local Similarity  83.6%;  Pred. No. 3.7e-44;
Matches  424;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps     0;

Qy      40 GCCCTGAACAGCCCAGCCCGGGCCGGGCATGCGAGGCATCCGGGAGGCGACTTCCAGTGC  99
           ||||  ||||| ||| ||||||||||||| |||| |||||||| |||||||||||||||
Db     682 GCGCTCAACAGCCCCCTGTCAGCGNCATGCGGGGGCATCCGGGGNCCGACTTNCAGTGC  623

Qy     100 TTCCAGCAGGCGGCGCGCGCCGGGCTGGCCGGACACTTTCCGGGCCTTCTGTCGTCGCGG  159
           ||||||||||||||||  || |||||| |||||||||||||||||| ||||||||| ||
Db     622 TCCAAGCAGGCGCGGNCCGTGGGCGTGGCCTGGGGGGCACTTTCCGGGCCTTCTGTCCTCGCGC  563

Qy     160 CTGCAGGACCTTCTACAGCATCGTGCGCCGGCCGACCCGGCACCGGGGTGCCCGTCGTCAAC  219
           ||||||||||||  ||| |||||||||||||||  |||| ||||||||||||| ||||||
Db     562 CTGCAGGACCTGCACAGACCATCGTGCGCCGCCGGTGCCGACCGCGCAGCCGTGCCCATCGTCAAC  503

Qy     220 CTCAGGGACAGAGTGCTCTTCCCCAGCTGGGAGGCCTTATTCCTCGGGCTCCCAGGGGCCAG  279
           ||||| ||||| |||||||  |||||| ||||||||||| |||||| |||| |||||||
Db     502 CTCAAGACGAGACGAGCTGCTGTTTCCCAGCTGGAGGCTCTGTTCTCAGGCTCTGAGGGTCCG  443

Qy     280 CTGAAGCCCGGGGGCCGCATCTTCTTTCGACGGCAGAGATGTCCTGCAGCACCCGCC  339
           |||||||| ||||||  ||||| |||  |||||||||||||| |||||||||||||
Db     442 CTGAAGCCCGGGGGCACGCATCTTCTTTGACGGCAAGGACGTCCTGAGGCACCCCACC  383

Qy     340 TGGCCCCGGAAGAGAGGCTGTGGCACGGGTCCGCAGGCCCCAGCGGGCCGGCCGACCAGC  399
           || | ||||||||| ||||||||| ||  ||||||  |||||||||||||| ||||||
Db     382 TGGCCCCGAGAAGAGAGGCTGTGGCCATGGCTCCGGACCTGGCGCAGGCCGCCCAGCGAGAGC  323

Qy     400 TACTGCGGACAGCGTGGCGGACGAGGCCCCGGGCACGAGGCCGCGGCGGGTCGTCGGCTG  459
           |||  || ||||| ||||| ||||||||| ||| ||||||| ||||||||||||||
Db     322 TACTGTGAGACGTGGGGGACCAGGAGGCTGTGGCATGGCTCCGGACCTGGCGCAGGCCGCCCAGCTG  263

Qy     460 GCGGGGCAGGTGCTGGAGCAGAGGAGGCCGGGCGAGCTGCCGGGAGTGCTGCTTCTGC  519
           ||||||||||| || ||||||||||||||||| ||| |||||| ||||||||||||
Db     262 GGGGGCAGGGTCCTGGGCAGAGTGCCGCAGAGTGCGCCAGCTGCCATCACGCCTACATCGTGCTCTGC  203

Qy     520 ATCGAGAACAGCGTCATGACCTCCTTC  546
           |||||||||||||||||||||||||
Db     202 ATTGAGAACAGCTTCATGACTGACTCC  176
```

RESULT 13
BF385854

```
LOCUS       BF385854                    843 bp   mRNA    linear   EST 27-NOV-2000
DEFINITION  602046021F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4195660 5',
            mRNA sequence.
ACCESSION   BF385854
VERSION     BF385854.1  GI:11367159
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 843)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9531 row: i  column: 05
            High quality sequence stop: 761.

FEATURES             Location/Qualifiers
     source          1..843
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4195660"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Li9"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match          67.7%;   Score 375.8; DB 10;  Length 843;
Best Local Similarity  79.8%;  Pred. No. 5e-44;
Matches  443;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps     0;

Qy       1 CACACCCACCAGGACTTCCAGCTGGTGCTGCACCTGGTGCCCTGGACAGCCCGCAGCCG  60
           |||  ||||||||||||||| | |||| ||||||||||| |||| | ||||||||||
Db     171 CATACTCATCAGGACTTTCAGCCAGTGCTCCACCTGGTGGCACTGAACACCCCCCTGTCT  230

Qy      61 GGGGGCATGCGAGGCATCCGGGAGGCGACTTCCAGTGCTTCCAGCAGGCGGCGCGCGCCG  120
           |||||||||||||| ||||||||| |||||||||||||| ||||||||||| ||| ||
Db     231 GGAGGCATGCGTGGTATCCGTGGAGCAGATTTCCAGTGCTTCCAGCAAGCCCGAGCCGTG  290

Qy     121 GGGCTGGCCGGACACCTTCCGTGCGTCGGCGCCTTCCTCGTGTCGCGGCTCAGCAGCATC  180
           |||||||| |||||||| ||||||| ||||  || ||||||||||||||||  ||||||
Db     291 GGGCTGTCGGGACACCTTCCGGGGCTTCCTTAGCGTGCCGGCTGCCGCAGGATCTCTATAGCATC  350

Qy     181 GTGCGCCGCGCCGACCGGCACCGGGGGTCGTGCCCATCGTGTCGCCATCGTCAACCTGTCTGTCTTC  240
           ||||||||||||||||||||| ||||  ||||| |||| | ||| ||||||||||| |
Db     351 GTGCGCGGTGCTGACCGGGGGGTCGTGCCCATCGTGTCCCATCGTCAACCTGAAGACGAGGTGCTCTTC  410

Qy     241 CCCAGCTGGGAGGCCTTATTCCTCGGGCTCCCAGGGGCCAGTGAAGCGGGCCCGCATC  300
           | ||||||| |||||||| |||||| |||| ||||||| |||||| ||||||||| |
Db     411 CCCAGCTGGGACTCCCTGTTTTCTGGCTCCCAGGGTCACTGACAACCCCGGGGGCCCGATC  470

Qy     301 TTCTCTTTCGACGGCAGAGATGTCCTGACGGCAAGGACGTCCTGAGGCACCCGTGG  360
           |||| |||| |||||||||||||||||||||| |||||||||| |||||||||||||
Db     471 TTTTCTTTCGACGGCAGAGATGTCCTGAGACACCCCGACGACAGCAGTCGGCAGAGGGTATGG  530

Qy     361 ATCGAGAACAGCGTCATGACCTCCTTC  420
           ||||||||| ||||||||||||||||
Db     531 CACGGCTTCGACCCCAGTGGGCGGAGGCTCATGAGGAGTTACTGTGAGACATGGCGAACT  590

Qy     421 GAGGCCCCGGGCACGGCGCACCGGGGCCAGGCCGGTGTCGCTGGCGGGCAGGCATGGCAGGCGGCTGCTCGGAGCAG  480
```

```
Db    203 GGAGGCATGCGTGGTATCGTGGAGCAGATTTCCAGTGCTTCCAGCAAGCCCGAGCCGTG 262
           |||||| ||||| ||| | ||||||  ||||||||||||||||||||| ||||||| |||
Qy    121 GGGCTGGCCGGCACCTTCCGGGCCTTCGTCGTCGGCGGCTCGAGGACCTTCACAGCATC 180
           |||||| ||||| ||| | ||||||  |||||||||||||| ||||| ||||||| ||||
Db    263 GGGCTGTCGGGCACCTTCCGGGCTTTCCGTCGTCCTCCAGGCTGCAGGATCTCTATAGCATC 322

Qy    181 GTGCGGCCGCGGACCGGCACCGGGCGTGCCGTCGTCGCTCGTCGAGAGGTGCTCTTC 240
           |||||||||| | |||||||||| |||||||||||||||||||||||||||| | |||
Db    323 GTGCCGGCCGTGCTGACCGGGGGGTCGTGCCCATCGTCGTCAACCTGAAGGACGAGGTGCTATCT 382

Qy    241 CCCAGCTGGGAGGGCCTTATTCTCCGGGCTCCGAGGGGCTCCGAGGGCCAGCTGAAGCCCGGCGGGCCCCCATC 300
           |||||| ||| | ||| |||| ||||||| |||||||||||| ||| ||| |||| |||||||| ||||| ||||
Db    383 CCCAGCTGGGACTCCCTGTTTTCTGGGCTCCGCCAGGGTCAACTGCAACTGCAACCCGGCGGGCCCGGCATC 442

Qy    301 TTCTCTTCGACGGCAGAGATGTCTCGACGATGTCTGCAGCACCCGCGCTGGCCCGGCGTGTGG 360
           |||||||||||||||| |||| || || |||||| ||||||||| |||||||||| |||||||
Db    443 TTTTCTTTGACGGCAGAGATGTCCGAGACACACCCCAGCCTGCCGCCGCAGAAGCGTATGG 502

Qy    361 CACGGCTCCACCCCAGCGGCGCGGCCTACTGCGACACAGCTACTGCGAGACGTTGGCGACG 420
           ||||||||||||||||| ||||||||||||  || ||| || ||||||| ||||||| |||
Db    503 CACGGCTCGGACCCCAGTGGCGGGAGGGTGTCGTCGTCGGGCCAGGTGCTGTCTGGAGCAG 562

Qy    421 GAGGGCCCGGGGCCACCGGGCAGGCGTGTCGTCGGCTGCTGTGGAGCAG 480
           |||||||| |||||||||  ||||| ||||||||| ||||||| ||||||
Db    563 GAAACTACTGGGGCTACAGGTCAGGCGTCAGGCTCCTGGAACAG 622

Qy    481 GAGGCCGGCGAGCTGCCACGCCTTCGTGGTGCTCTGCATGGAGAACACAGGGTCATGACC 540
           |||||||||||||||||||  || |||||| ||||||||||||||| |||||||||||||
Db    623 AAAGTTGCGAGCTGCCACACAGCTACATCGTCCTGTGCCATTGAGAATAGCTTCATGACC 682

Qy    541 TCCTTCTCCAAGTAG 555
           |||||||||||||||
Db    683 TCTTTCTCCAAATAG 697
```

RESULT 15

BC062931

LOCUS   BC062931   4192 bp   mRNA   linear   HTC 17-DEC-2003
DEFINITION   Mus musculus cDNA clone IMAGE:30354418, containing frame-shift
                 errors.
ACCESSION   BC062931
VERSION   BC062931.1 GI:38614356
KEYWORDS   HTC.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
                 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 4192)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
                 Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
                 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
                 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
                 Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
                 Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
                 Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
                 Carninci,P., Prange,C., Raha,S.S. Loquellano,N.A., Peters,G.J.,
                 Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
                 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
                 Worley,K.C. Hale,S., Garcia,A.M. Gay,L.J., Hulyk,S.W.,
                 Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
                 Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
                 Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
                 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
                 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
                 Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
                 Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
                 human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  PUBMED   12477932
REFERENCE   2 (bases 1 to 4192)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission

```
Db    591 GAAACTACTGGGGCTACAGTCAGGCCTCCTCCCCGTGTCAGGCAGGCTCCTGGAACAG 650
           |||| | || || || || | |||  ||  ||||||||||||||||||| |||||| |||
Qy    481 GAGGCCGGCGAGCTGCCGCACGCCTTCGTCCTGGTGTCGCATCGAGAACACAGGGTCATGACC 540
           ||||||||||||||||||| ||||  ||||| || |||||||||| |||||||||| |||||
Db    651 AAAGTGCAGCTGCCACCAAACAGCAGTACATCGTCCTGTGCCTGTGCATTGAGAATAGCTTCATGACC 710

Qy    541 TCCTTCTCCAAGTAG 555
           |||||||||||||||
Db    711 TCTTTCTCCAAATAG 725
```

RESULT 14

BI412588

LOCUS   BI412588   874 bp   mRNA   linear   EST 14-AUG-2001
DEFINITION   60290468F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5146409 5',
                 mRNA sequence.
ACCESSION   BI412588
VERSION   BI412588.1 GI:15173511
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
                 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 874)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
                 Email: cgapbs-r@mail.nih.gov
                 Tissue Procurement: Gilbert Smith, Ph.D.
                 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
                 Bonaldo, Ph.D.
                 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                 DNA Sequencing by:Incyte Genomics, Inc.
                 Clone distribution: NCI-CGAP clone distribution information can be
                 found through the I.M.A.G.E. Consortium/LLNL at:
                 http://image.llnl.gov
                 Plate: LLAM1361 row: g column: 18
                 High quality sequence start: 23
                 High quality sequence stop: 808.
FEATURES   Location/Qualifiers
     source   1..874
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="Czech II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5146409"
                     /tissue_type="pooled lung tumors"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NCI_CGAP_Lu33"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
                     strand cDNA was prepared from mRNA obtained from pooled
                     lung tumors with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGGCGGCCGCCTCTGTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN
Query Match   67.7%; Score 375.8; DB 12; Length 874;
Best Local Similarity   79.8%; Pred. No. 5e-44;
Matches   443; Conservative   0; Mismatches 112; Indels   0; Gaps   0;

```
Qy      1 CACACCCACCAGGACTTCCAGCTTGCACCTGTGCACGTTGGCCCTGAACAGCCCGCAGCCG 60
          |||||| | ||||| ||||||| || |||||||||||||| |||||||| |||||| ||||
Db    143 CATACTCATCAGGACTTTCAGCCAGTGCTGCCACCTGGTGGCACTGAACACACCCCCTGTCT 202

Qy     61 GGGCGGCCATGCGAGGCATCCGGGGAGGCCGGACTTCCAGGTGCTTCCAGCAGGGCGCGCCCG 120
          |||||||||| || |||  ||| |||||  |||  ||||||||||||| || | ||||  |||
```

JOURNAL Submitted (26-NOV-2003) National Institutes of Health, Mammalian
  Gene Collection (MGC), Cancer Genomics Office, National Cancer
  Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
  USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: cgapbs-r@mail.nih.gov
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: Dr. Jim Lin, University of Iowa
  cDNA Library Preparation: M. Bento Soares, University of Iowa
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
  Thomas L. Casavant.
  Web site: http://genome.uiowa.edu
  Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
  Bonaldo,M.F., Akabogu,I., Bair,T., Bair,J., Crouch,K., Davis,A.,
  Fishler,K., Keppel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K.,
  Scheetz,T., Smith,C., Snir,E., Tack,D., Trout,K., Walters,J.,
  Casavant,T., Soares,M.B.

  Clone distribution: MGC clone distribution information can be found
  through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
  Series: Plate: Row: Column: 0
  This clone was selected for full length sequencing because it
  passed the following selection criteria: matched mRNA gi: 13937350
  This clone has the following problem: frame shifted.
FEATURES Location/Qualifiers
 source 1..4192
  /organism="Mus musculus"
  /mol_type="mRNA"
  /strain="C57BL/6"
  /db_xref="taxon:10090"
  /clone="IMAGE:30354418"
  /tissue_type="Brain, mouse, 13.5,14.5,16.5,17.5 dpc"
  /clone_lib="NIH_BMAP_FY0"
  /lab_host="DH10B"
  /note="Vector: pYX-ASC"

ORIGIN

Query Match 67.7%; Score 375.8; DB 11; Length 4192;
Best Local Similarity 79.8%; Pred. No. 6.6e-44;
Matches 443; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```
Qy     1 CACACCCACCAGGACTTCCAGCTGGTGTGCTGCACCTGGTGGCCCTGAACAGCCCGCAGCCG   60
          |||||||| ||| ||||||| ||||  ||||||| ||||||||||||| || ||||||||||
Db  3442 CATACTCATCAGGACTTTCAGCCAGTGCTCACCTGGTGGCACTGAACACCCCCCTGTCT 3501

Qy    61 GGCGGCATCGGAGGCATCCGGGAGGCAGACTTCCAGTGCTTCCAGCAGGCGCGCCGCCGCG  120
          ||| |||| ||| |||| |||||  || | ||||| |||| |||| || || | | | |||
Db  3502 GGAGGCATGCGTGTATCCGTGGACAGAGATTTCCAGTGCTTCCAGCAGCCCGAGCCGTG 3561

Qy   121 GGGCTGGCCGGCACCTTCCGGGGCCTTCCTGTGCGCGGCTGCGAGGACCTCTACAGCATC  180
          |||||||| | ||||| ||||| ||  |||||  ||||| || ||| || ||| ||||
Db  3562 GGGCTGTCGGGCACCTTCCGGGGCTTTCCTGTCCTCTCAGGCTGCGAGGATCTCTATAGCATC 3621

Qy   181 GTGGGCCGCGCGCCACCGCACCGGGGTGCCCGTCGTCAACCTCAGGGACGAGGTGCTCTTC  240
          |||||||| ||||||| || | |||||||| |||||||| ||||| || ||||||| |
Db  3622 GTGCCCGCCGTCTGACGCGGGGGGTCTGTGCCATGTCAACCTGAAGGACGAGGTGTATCT 3681

Qy   241 CCCAGCTGGGAGGCCTTATTCTGGGCTCGGAGCCCAGCTGAAGGCCCGGGCCCGCCATC  300
          || |||||||||||||  |||||| ||| ||||| ||||| ||| ||| || ||||||
Db  3682 CCCAGCTGGGACTCCTGTTTTCTGGGTCAACTGCAACCCGGGGCCCGGGACCCGCCATC 3741

Qy   301 TTCTCTTTCGACGGCAGAGATGTCCTGAGCACCCCGCTGGCCCCGAAGAGCGTGTGG  360
          |||||||| |||||| ||||||| |||||||||||  |||| |||||||| ||||||
Db  3742 TTTTCTTTTGACGGCAGAGATGTCCTGAGCACCCCGCAGAGAGCGTATCG 3801

Qy   361 CACGGCTCCGACCCCAGCGGCGCCGCCGCCTGACCGACACTGCCGAGACGTGGCGGACG  420
          ||||||||| |||||| ||| ||| | || | || || || || || |||| || || ||
Db  3802 CACGGCTCGGACCCCAGTGGCGGAGGCTGATGGAGAGTTACTGTGAGACCATGGCCGAACT 3861

Qy   421 GAGGCCCCGGCCCGGCCACCGGGCCAGGCCGTCGTCCTGTGCTCGTGGCGGCCAGGCAG  480
          || | |||| |||| ||| |||| ||||| || |||||||||||||||| ||||||| ||
Db  3862 GAAACTACTGGGGCTACAGGTCAGGCCTCCTCCGCCTGCTGGTCGTGTCCTGGAACAG 3921
```

(right column)

```
Qy   481 GAGGCCCGGAGCTGCGCCACCGCCCTTCGTGTGTGCTTCTGCATCGAGAACAGCGTCATGACC  540
          |||||||| |||  ||||| | | |  |||||||||||| |||| |||  | || |||||||
Db  3922 AAAGCTGCGAGCTGCCACAACAGCTACATCCGTCCTGCATTGAGAATAGCTTCATGACC 3981

Qy   541 TCCTTCTCCAAGTAG  555
          || ||||| ||| |||
Db  3982 TCTTTCTCCAAATAG  3996
```

Search completed: March 29, 2004, 14:52:46
Job time : 1706.7 secs

```
OM protein - protein search, using sw model

                   GenCore version 5.1.6
              Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:    March 26, 2004, 13:28:41 ; Search time 45.7778 Seconds
                                      (without alignments)
                                      1135.676 Million cell updates/sec

Title:          US-09-938-391-4
Perfect score:  966
Sequence:       1 HTHQDFQLIVLHLVALNSPQP.........CRHAFVVLCIENSVMTSFSK 184

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:    1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    A_Geneseq_29Jan04:*
           1: geneseqp1980s:*
           2: geneseqp1990s:*
           3: geneseqp2000s:*
           4: geneseqp2001s:*
           5: geneseqp2002s:*
           6: geneseqp2003as:*
           7: geneseqp2003bs:*
           8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                          SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	966	100.0	966	5	AAO17430	Canine en
2	966	100.0	230	5	AAO17429	Canine pr
3	959	99.3	184	3	AAY70265	Canine an
4	850	88.0	184	5	ABG31794	Murine en
5	835	86.4	181	4	AAU00898	Human End
6	835	86.4	182	3	AAY59622	Human end
7	835	86.4	182	3	AAY94323	Human end
8	835	86.4	182	4	AAB28399	Human end
9	835	86.4	182	4	AAU00897	Human End
10	835	86.4	182	2	AAU77951	Amino aci
11	835	86.4	183	2	AAY02113	SEQ ID 76
12	835	86.4	183	2	AAY08693	Human end
13	835	86.4	183	3	AAY70252	Human ang
14	835	86.4	183	3	AAY90771	Human ang
15	835	86.4	183	3	AAB16451	Human end
16	835	86.4	183	3	AAB30493	Amino aci
17	835	86.4	183	4	AAB49379	Human end
18	835	86.4	183	4	AAU00896	Human End
19	835	86.4	183	5	ABB79901	Human end
20	835	86.4	183	5	AAM49503	Human end
21	835	86.4	183	5	AAM48895	Human end
22	835	86.4	183	5	AAU97132	Human end
23	835	86.4	183	6	AAG79753	Human end
24	835	86.4	195	3	AAW90874	Human HMW
25	835	86.4	216	3	AAB30495	Amino aci
26	835	86.4	275	5	AAU76689	Synthetic
27	835	86.4	310	5	AAU76688	Human col
28	835	86.4	513	5	ABG73586	Human End
29	835	86.4	682	5	ABP41878	Human ova
30	835	86.4	684	2	AAW26327	Human alp
31	835	86.4	684	5	AAY25113	Human alp
32	835	86.4	684	5	AAO17357	Human col
33	835	86.4	1301	2	AAW92296	Human alp
34	835	86.4	1336	2	AAY08694	Human col
35	835	86.4	1336	6	ABP96308	Human end
36	835	86.4	1516	5	ABB83471	Human col
37	835	86.4	1516	4	ABP68617	Human pan
38	834	86.3	180	5	AAU00899	Human End
39	834	86.3	193	3	AAW90877	Human HMW
40	831	86.0	183	4	AAB49810	Human end
41	831	86.0	184	4	AAB49380	Murine en
42	831	86.0	191	3	AAB28398	Murine en
43	831	86.0	191	5	AAU77950	Amino aci
44	831	86.0	207	5	ABB79902	Mouse end
45	828	85.7	184	2	AAY08689	Murine en

```
                          ALIGNMENTS

RESULT 1
AAO17430
ID   AAO17430 standard; protein; 184 AA.
XX
AC   AAO17430;
XX
DT   19-JUL-2002  (first entry)
XX
DE   Canine endostatin.
XX
KW   Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;
KW   psoriasis; rheumatoid arthritis; retinopathy; macular degeneration;
KW   corneal graft rejection; neovascular glaucoma; retrolental fibrop-asia;
KW   rubeosis; Osler-Webber Syndrome; myocardial angiogenesis;
KW   plaque neovascularisation; telangiectasia; haemophiliac joints;
KW   angiofibroma; wound granulation; coronary collateral;
KW   cerebral collateral; arteriovenous malformation;
KW   ischaemic limb angiogenesis; diabetic neovascularisation; fracture;
KW   cytostatic; antiarthritic; antipsoriatic; antidiabetic; ophthalmological;
KW   gynaecological.
XX
OS   Canis familiaris.
XX
PN   EP1191036-A2.
XX
PD   27-MAR-2002.
XX
PF   24-AUG-2001; 2001EP-00307224.
XX
PR   25-AUG-2000; 2000US-0227924P.
XX
PA   (PFIZ ) PFIZER PROD INC.
XX
PI   Sheppard MG, Tong X;
XX
DR   WPI; 2002-354068/39.
DR   N-PSDB; AAL46063.
XX
PT   An isolated nucleic acid molecule for the treatment of angiogenesis-
PT   related disorder, such as cancers or diabetic retinopathy, encodes an
PT   endostatin protein.
XX
PS   Claim 14; Fig 5; 56pp; English.
XX
CC   The present invention provides the protein and coding sequences of canine
CC   pro-endostatin and endostatin. The sequences can be used in the treatment
CC   and diagnosis of angiogenesis related disorders, including cancer,
CC   rheumatoid arthritis, psoriasis, retinopathy, macular degeneration,
```

XX
CC   The present invention provides the protein and coding sequences of canine
CC   pro-endostatin and endostatin. The sequences can be used in the treatment
CC   and diagnosis of angiogenesis related disorders, including cancer,
CC   rheumatoid arthritis, psoriasis, retinopathy, macular degeneration,
CC   corneal graft rejection, neovascular glaucoma, retrolental fibroplasia,
CC   rubeosis, Osler-Webber Syndrome, myocardial angiogenesis, plaque
CC   neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,
CC   wound granulation, coronary collaterals, cerebral collaterals,
CC   arteriovenous malformations, ischaemic limb angiogenesis, diabetic
CC   neovascularisation, and fractures. The present sequence is the canine pro
CC   -endostatin protein sequence
XX
SQ   Sequence 230 AA;

Query Match   100.0%;   Score 966; DB 5; Length 230;
Best Local Similarity 100.0%;   Pred. No. 7.8e-107;
Matches 184; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   47 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 106

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  107 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 166

Qy  121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  167 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 226

Qy  181 SFSK 184
          ||||
Db  227 SFSK 230
```

RESULT 3
AAY70265
ID   AAY70265 standard; protein; 184 AA.
XX
AC   AAY70265;
XX
DT   06-JUN-2000 (first entry)
XX
DE   Canine angiogenesis inhibitor, endostatin.
XX
KW   Canine; immunoglobulin Fc fragment; endostatin; immunofusin;
KW   angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;
KW   antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;
KW   vasotropic; vulnerary; treatment; antiarteriosclerosis; tumour;
KW   metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
KW   ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW   myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW   wound granulation; keloid scar; gene therapy.
XX
OS   Canis familiaris.
XX
PN   WO200011033-A2.
XX
PD   02-MAR-2000.
XX
PF   25-AUG-1999;   99WO-US019329.
XX
PR   25-AUG-1998;   98US-0097883P.
XX
PA   (LEXI-) LEXINGEN PHARM. CORP.
XX
PI   Lo K, Li Y, Gillies SD;
XX
DR   WPI; 2000-237616/20.
DR   N-PSDB; AAZ51309.
XX

CC   corneal graft rejection, neovascular glaucoma, retrolental fibroplasia,
CC   rubeosis, Osler-Webber Syndrome, myocardial angiogenesis, plaque
CC   neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,
CC   wound granulation, coronary collaterals, cerebral collaterals,
CC   arteriovenous malformations, ischaemic limb angiogenesis, diabetic
CC   neovascularisation, and fractures. The present sequence is the canine pro
CC   -endostatin protein sequence
XX
SQ   Sequence 184 AA;

Query Match   100.0%;   Score 966; DB 5; Length 184;
Best Local Similarity 100.0%;   Pred. No. 5.6e-107;
Matches 184; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI  60

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120

Qy  121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180

Qy  181 SFSK 184
          ||||
Db  181 SFSK 184
```

RESULT 2
AAO17429
ID   AAO17429 standard; protein; 230 AA.
XX
AC   AAO17429;
XX
DT   19-JUL-2002 (first entry)
XX
DE   Canine pro-endostatin.
XX
KW   Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;
KW   psoriasis; rheumatoid arthritis; retinopathy; macular degeneration;
KW   corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
KW   rubeosis; Osler-Webber Syndrome; myocardial angiogenesis;
KW   plaque neovascularisation; telangiectasia; haemophiliac joints;
KW   angiofibroma; wound granulation; coronary collateral;
KW   cerebral collateral; arteriovenous malformation;
KW   ischaemic limb angiogenesis; diabetic neovascularisation; fracture;
KW   cytostatic; antiarthritic; antipsoriatic; antidiabetic; ophthalmological;
KW   gynaecological.
XX
OS   Canis familiaris.
XX
PN   EP1191036-A2.
XX
PD   27-MAR-2002.
XX
PF   24-AUG-2001; 2001EP-00307224.
XX
PR   25-AUG-2000; 2000US-0227924P.
XX
PA   (PFIZ) PFIZER PROD INC.
XX
PI   Sheppard MG, Tong X;
XX
DR   WPI; 2002-354068/39.
DR   N-PSDB; AAL46062.
XX
PT   An isolated nucleic acid molecule for the treatment of angiogenesis-
PT   related disorder, such as cancers or diabetic retinopathy, encodes an
PT   endostatin protein.
XX

PT   Novel fusion protein of angiostatin or endostatin and an immunoglobulin
PT   FC region, useful for treating conditions mediated by angiogenesis, such
PT   as rheumatoid arthritis, tumors and macular degeneration.
XX
PS   Example 8; Page 59-60; 68pp; English.
XX
CC   The patent discloses a DNA molecule encoding a fusion protein comprising
CC   a signal sequence, an immunoglobulin Fc region, and an angiogenesis
CC   inhibitor selected from angiostatin, endostatin, a plasminogen fragment
CC   having angiostatin activity, a collagen XVIII fragment having endostatin
CC   activity, or combinations of them. The fusion protein (immunofusin) is
CC   used to inhibit angiogenesis and to treat diseases or conditions mediated
CC   by angiogenesis. Conditions that may be treated include solid tumours,
CC   blood born tumours, tumour metastasis, benign tumours including
CC   haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic
CC   granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
CC   e.g. diabetic retinopathy, retinopathy of prematurity, macular
CC   degeneration, corneal graft rejection, neovascular glaucoma, retrolental
CC   fibroplasia, rubeosis and Osler-Webber syndrome; myocardial angiogenesis,
CC   plaque neovascularisation, telangiectasia, haemophiliac joints'
CC   angiofibroma, wound granulation, and excessive or abnormal stimulation of
CC   endothelial cells, intestinal cells, atherosclerosis, scleroderma and
CC   hypertrophic scars, i.e. keloid scars. The DNA constructs may be used in
CC   gene therapy. The present sequence is a canine endostatin used in the
CC   construction of immunofusin containing canine immunoglobulin Fc fragment
XX
SQ   Sequence 184 AA;

Query Match   99.3%; Score 959; DB 3; Length 184;
Best Local Similarity   99.5%; Pred. No. 3.9e-106;
Matches 183; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 HTHQDFQPVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120

Qy   121 HGSDPSGRRLTDSYCETWRTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 HGSDPSGRRLTDSYCETWRTEAPATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180

Qy   181 SFSK 184
          ||||
Db   181 SFSK 184

RESULT 4
ABG31794
ID   ABG31794 standard; protein; 184 AA.
XX
AC   ABG31794;
XX
DT   05-DEC-2002 (first entry)
XX
DE   Murine endostatin polypeptide.
XX
KW   Mouse; endostatin; tumour; cancer; metastasis; cytostatic;
KW   antiangiogenic.
XX
OS   Mus sp.
XX
PN   WO20026457-A2.
XX
PD   06-SEP-2002.
XX
PF   27-FEB-2002; 2002WO-IT000119.
XX
PR   27-FEB-2001; 2001IT-MI000394.
XX
PA   (UYMI-) UNIV MILANO.

XX
PI   Chillemi F, Vicentinie LMT, Francescato P;
XX
DR   WPI; 2002-698655/75.
XX
PT   New peptide useful for the preparation of medicaments with antiangiogenic
PT   activity that may be used in treating tumors or metastases, comprises a
PT   sequence corresponding to fragments of human endostatin.
XX
PS   Disclosure; Fig 1; 24pp; English.
XX
CC   The invention relates to peptide comprising 20-50 amino acids with
CC   sequences corresponding to the human endostatin polypeptide sequence, its
CC   salt or non-toxic derivative. The peptides are useful in the preparation
CC   of medicaments with antiangiogenic activity which may be useful in
CC   treating tumours or metastases. This sequence represents a murine
CC   endostatin polypeptide
XX
SQ   Sequence 184 AA;

Query Match   88.0%; Score 850; DB 5; Length 184;
Best Local Similarity   85.3%; Pred. No. 4.4e-93;
Matches 157; Conservative   14; Mismatches   13; Indels   0; Gaps   0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
          |:|:|| ||||||||||| :|||||||||||||||||||| ||||||||||||||||||||
Db   1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy   61 VRRADRTGVPVVNLRDBVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
          |||||| :| ||| |: ||||||||||||||| ||||||||||| ||| :|| ||:||||
Db   61 VRRADRAAVPIVNLKDELLFFSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy   121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
          |||||:||||||:||||||||||: |||||||| |||| ::|||| :||:||||||| ||
Db   121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGGQSAASCHHAYIVLCIENSFMT 180

Qy   181 SFSK 184
          :|||
Db   181 AFSK 184

RESULT 5
AAU00898
ID   AAU00898 standard; protein; 181 AA.
XX
AC   AAU00898;
XX
DT   04-JUL-2001 (first entry)
XX
DE   Human Endostatin(TM) C-terminus minus 2 protein.
XX
KW   Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;
KW   blood borne tumour; leukaemia; tumour metastasis; benign tumour;
KW   haemangioma; acoustic neuroma; neurofibroma; trachoma; rubeosis;
KW   pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;
KW   ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW   retinopathy of prematurity; macular corneal graft rejection;
KW   neovascular glaucoma; retrolental fibroplasia; Osler-Webber Syndrome;
KW   myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW   haemophiliac joint; angiofibroma; wound granulation; variant;
KW   C-terminus minus 2 protein.
XX
OS   Homo sapiens.
XX
PN   WO200119989-A2.
XX
PD   22-MAR-2001.
XX
PF   14-SEP-2000; 2000WO-US025166.
XX
PR   14-SEP-1999; 99US-015369BP.
XX
PA   (ENTR-) ENTREMED INC.

```
XX
PI   Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ;
PI   Bermejo LL, Mistry FR, Shepard SR, Schrimsher JL;
XX
DR   WPI; 2001-244802/25.
DR   N-PSDB; AAS00867.
XX
PT   Producing Endostatin protein for treating angiogenesis mediated diseases
PT   such as solid tumors, comprises recombinantly producing the protein using
PT   an expression system, and recovering and purifying the protein.
XX
PS   Claim 5; Page 30; 67pp; English.
XX
CC   The sequence represents Human Endostatin(TM) C-terminus minus 2 protein,
CC   a natural variant lacking the C-terminal 2 amino acids of Endostatin(TM)
CC   recovered from fermentations of Pichia pastoris cultures harbouring a
CC   expression plasmid containing the Endostatin(TM) DNA sequence given in
CC   AAS00867. The new method of the invention is useful for producing,
CC   recovering and purifying Endostatin (TM) from biological sources, such as
CC   biological fluids, tissues, cells, culture media, and fermentation media.
CC   Endostatin(TM) is useful for treating angiogenesis mediated diseases such
CC   as solid tumours, blood borne tumours, leukaenias, tumour metastases,
CC   benign tumours, e.g. haemangioma, acoustic neuroma, neurofibromas,
CC   trachomas, and pyogenic granulomas, rheumatoid arthritis, psoriasis,
CC   ocular angiogenic diseases, e.g. diabetic retinopathy, retinopathy of
CC   prematurity, macular degeneration, corneal graft rejection, neovascular
CC   glaucoma, colon cancer, retrolental fibroplasia, rubeosis, Osler-Webber
CC   Syndrome, myocardial angiogenesis, plaque neovascularisation,
CC   telangiectasia, haemophiliac joints, angiofibroma, and wound granulation.
CC   Endostatin(TM) is also useful for treating diseases of excessive or
CC   abnormal stimulation of endothelial cells such as intestinal adhesions,
CC   atherosclerosis, scleroderma and hypertrophic scars. Higher yields of
CC   more purified, and biologically active Endostatin(TM) are obtained by the
CC   new method. Endostatin(TM) can be stored in buffers for extended periods
CC   of time, and also subjected to lyophilisation, while preserving
CC   biological activity. Centrifugation of broth from fermentation steps in
CC   production is avoided, preventing unwanted potential cellular lysis and
CC   contamination with additional proteins, pigments, enzymes and other
CC   cellular chemicals and debris
XX
SQ   Sequence 181 AA;

Query Match          86.4%;  Score 835;  DB 4;  Length 181;
Best Local Similarity 85.1%;  Pred. No. 2.7e-91;
Matches  154;   Conservative  14;  Mismatches  13;   Indels   0;   Gaps   0;

Qy     1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSSRLQDLYSI 60
           |:|:||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSSRLQDLYSI 60

Qy    61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSPDGRDVLQHPAWPRKSVW 120
          |||||  |:|| :|: :|||| :||||||||:| ||:||||||||:||||:| :|::|
Db    61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy   121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
          |||||| || ||:|||||||||:|||||||||::|||||||||  |:||||||||||
Db   121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy   181 S 181
           :
Db   181 A 181

RESULT 6
AAY59622
ID   AAY59622 standard; protein; 182 AA.
XX
AC   AAY59622;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Human endostatin protein fragment.
XX
```

```
KW   Endostatin; scatter factor activity; human; tubulogenesis; psoriasis;
KW   metastatic cancer; tumourigenesis; ocular angiogenic disease;
KW   rheumatoid arthritis; Osler-Webber syndrome; telangiectasia;
KW   haemophiliac joint; angiofibroma; wound granulation.
XX
OS   Homo sapiens.
XX
PN   WO9962944-A2.
XX
PD   09-DEC-1999.
XX
PF   03-JUN-1999;   99WO-US012278.
XX
PR   03-JUN-1998;   98US-0087890P.
PR   10-JUL-1998;   98US-0092393P.
PR   01-SEP-1998;   98US-0098790P.
XX
PA   (CHIL-) CHILDRENS MEDICAL CENT.
XX
PI   Javaherian K,  Folkman MJ;
XX
DR   WPI; 2000-072833/06.
XX
PT   New endostatin oligomers, used for treating e.g. tumors.
XX
PS   Disclosure; Page 6; 44pp; English.
XX
CC   This sequence is a fragment of the human endostatin protein. Endostatin
CC   is an approximately 20kD C-terminal globular domain of the collagen-like
CC   protein collagen XVIII. Protein oligomers consisting of more than one
CC   endostatin monomer have anti-tubulogenic effects and induce
CC   reorganization of the actin cytoskeleton. The oligomer has scatter factor
CC   activity. The oligomers induce the destruction of tubular lumens and
CC   elongation of cells, and inhibit tubulogenesis and tumourigenesis. The
CC   oligomers can also be used to treat metastatic cancers, tumours,
CC   rheumatoid arthritis, psoriasis, ocular angiogenic disease, Osler-Webber
CC   syndrome, plaque neovascularisation, telangiectasia, haemophiliac joints,
CC   angiofibroma and wound granulation. The oligomers can also be used to
CC   treat diseases that have angiogenesis as a pathological consequence e.g.
CC   ulcers. The endostatin oligomers can also be used to develop affinity
CC   columns for isolating antibodies or receptors. Passive antibody therapy
CC   using antibodies that specifically bind endostatin oligomers can be used
CC   to modulate morphogenic processes such as metastatic cancer as well as
CC   angiogenesis-dependent processes such as reproduction, development, wound
CC   healing, tissue repair, and angiogenesis-dependent diseases. Also,
CC   antisera directed to the Fab regions of endostatin oligomer antibodies
CC   can be administered to block the ability of endogenous endostatin
CC   oligomer antisera to bind endostatin oligomers
XX
SQ   Sequence 182 AA;

Query Match          86.4%;  Score 835;  DB 3;  Length 182;
Best Local Similarity 85.1%;  Pred. No. 2.7e-91;
Matches  154;   Conservative  14;  Mismatches  13;   Indels   0;   Gaps   0;

Qy     1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSSRLQDLYSI 60
           |:|:||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSSRLQDLYSI 60

Qy    61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSPDGRDVLQHPAWPRKSVW 120
          |||||  |:|| :|: :|||| :||||||||:| ||:||||||||:||||:| :|::|
Db    61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy   121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
          |||||| || ||:|||||||||:|||||||||::|||||||||  |:||||||||||
Db   121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy   181 S 181
           :
Db   181 A 181

RESULT 7
```

```
AAY94323
ID   AAY94323 standard; protein; 182 AA.
XX
AC   AAY94323;
XX
DT   11-AUG-2000  (first entry)
XX
DE   Human endostatin protein.
XX
KW   Human; endothelial cell proliferation inhibitor; collagen XVIII;
KW   angiogenesis inhibitor; anti-tumour; cytostatic; antipsoriatic;
KW   vasotropic; determatological; ophthalmological; vulnerary;
KW   antiarteriosclerotic; antidiabetic; haemostatic; contraceptive;
KW   ocular angiogenic disease; atherosclerosis; scleroderma;
KW   myocardial angiogenesis; telangiectasia; angiofibroma; wound granulation.
XX
OS   Homo sapiens.
XX
PN   WO200026368-A2.
XX
PD   11-MAY-2000.
XX
PF   01-NOV-1999;  99WO-US025605.
XX
PR   30-OCT-1998;  98US-0106343P.
PR   20-MAY-1999;  99US-00315889.
XX
PA   (CHIL-) CHILDRENS MEDICAL CENT.
XX
PI   O'reilly MS, Folkman MJ;
XX
DR   WPI; 2000-365617/31.
DR   N-PSDB; AAA27004.
XX
PT   Novel endostatin capable of inhibiting endothelial cell proliferation and
PT   angiogenesis, useful for treating angiogenesis-dependent cancers and as
PT   birth control agents.
XX
PS   Claim 2; Page 38; 68pp; English.
XX
CC   The present sequence is an endostatin protein which is the carboxy
CC   terminal protein of human collagen XVIII. Recombinant mouse endostatin
CC   (20 mg/kg) was administered subcutaneously to mice implanted with Lewis
CC   lung carcinomas. There was tumour mass regression non-detectable levels
CC   after 12 days of therapy due to the angiogenesis inhibitory activity of
CC   endostatin. Thus the protein is useful for treatment of angiogenesis-
CC   dependent cancers. The polynucleotide and polypeptide sequences of this
CC   endostatin are useful for treating and diagnosis of tumours, ocular
CC   angiogenic diseases, Osler-Webber syndrome, myocardial angiogenesis,
CC   plaque neovascularisation, telangiectasia, haemophiliac joints,
CC   angiofibroma and wound granulation, for treatment of diseases related to
CC   excessive or abnormal stimulation of endothelial cells e.g. intestinal
CC   adhesions, atherosclerosis, scleroderma. The protein may also be useful
CC   as a birth control agent by reducing or preventing uterine
CC   vascularisation. The gene for endostatin may be isolated from cells or
CC   tissue that express high levels of endostatin, eg. tumour cells, by
CC   generating cDNA from mRNA using reverse transcriptase and then amplifying
CC   the DNA sequence
XX
SQ   Sequence 182 AA;

Query Match             86.4%;  Score 835;  DB 3;  Length 182;
Best Local Similarity   85.1%;  Pred. No. 2.7e-91;
Matches  154;  Conservative  14;  Mismatches  13;  Indels  0;  Gaps  0;

Qy          1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
              |:|:|| ||||||||||||| |:||||||||||||  ||||||||||||||||||||||
Db          1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy         61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
              |||||  |||  ||| |:|||||||||||||:|||||||||||||:|||:||  |:||||
Db         61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy        121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
              ||||| ||||||:||||||||| |:|||||||||||||||||: ||:||  |||||| |
Db        121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGQSAASCHAYIVLCIENSFMT 180

Qy        181 S 181
              :
Db        181 A 181
```

```
RESULT 8
AAB28399
ID   AAB28399 standard; protein; 182 AA.
XX
AC   AAB28399;
XX
DT   19-FEB-2001  (first entry)
XX
DE   Human endostatin.
XX
KW   Human; endostatin; cytostatic; antiproliferative;
KW   vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;
KW   cancer; vascularised solid tumour.
XX
OS   Homo sapiens.
XX
PN   WO200064946-A2.
XX
PD   02-NOV-2000.
XX
PF   28-APR-2000;  2000WO-US011367.
XX
PR   28-APR-1999;  99US-0131432P.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Thorpe PE,  Brekken RA;
XX
DR   WPI; 2000-687317/67.
XX
PT   Immunogenic composition for the treatment and diagnosis of cancer
PT   comprises an anti-VEGF (vascular endothelial growth factor) antibody
PT   binding the same epitope as the monoclonal antibody ATCC PTA 1595.
XX
PS   Example 10; Page 291-292; 298pp; English.
XX
CC   The present invention relates to anti-Vascular Endothelial Growth Factor
CC   (VEGF) antibodies that bind to the same epitope as the monoclonal
CC   antibody ATCC PTA 1595 and which significantly inhibit VEGF binding to
CC   the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF
CC   receptor VEGFR1. The present sequence is human endostatin. Endostatin may
CC   be conjugated onto the anti-VEGF antibodies of the present invention. The
CC   anti-VEGF antibodies of the present invention are useful for the
CC   treatment and diagnosis of cancer, especially vascularised solid tumours
XX
SQ   Sequence 182 AA;

Query Match             86.4%;  Score 835;  DB 3;  Length 182;
Best Local Similarity   85.1%;  Pred. No. 2.7e-91;
Matches  154;  Conservative  14;  Mismatches  13;  Indels  0;  Gaps  0;

Qy          1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
              |:|:|| ||||||||||||| |:||||||||||||  ||||||||||||||||||||||
Db          1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy         61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
              |||||  |||  ||| |:|||||||||||||:|||||||||||||:|||:||  |:||||
Db         61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy        121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
              ||||| ||||||:||||||||| |:|||||||||||||||||: ||:||  |||||| |
Db        121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGQSAASCHAYIVLCIENSFMT 180

Qy        181 S 181
```

```
Db     181 A 181
      :

RESULT 9
AAU00897
ID   AAU00897 standard; protein; 182 AA.
XX
AC   AAU00897;
XX
DT   04-JUL-2001 (first entry)
XX
DE   Human Endostatin(TM) C-terminus minus 1 protein.
XX
KW   Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;
KW   blood borne tumour; leukaemia; tumour metastasis; benign tumour;
KW   haemangioma; acoustic neuroma; neurofibroma; trachoma; rubeosis;
KW   pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;
KW   ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW   retinopathy of prematurity; macular corneal graft rejection;
KW   neovascular glaucoma; retrolental fibroplasia; Osler-Webber Syndrome;
KW   myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW   haemophiliac joint; angiofibroma; wound granulation; variant;
KW   C-terminus minus 1 protein.
XX
OS   Homo sapiens.
XX
PN   WO200119989-A2.
XX
PD   22-MAR-2001.
XX
PF   14-SEP-2000; 2000WO-US025166.
XX
PR   14-SEP-1999;   99US-0153598P.
XX
PA   (ENTR-) ENTREMED INC.
XX
PI   Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J,   Boerner RJ;
PI   Bermejo LL, Mistry FR, Shepard SR,  Schrimsher JL;
XX
DR   WPI; 2001-244802/25.
DR   N-PSDB; AAS00897.
XX
PT   Producing Endostatin protein for treating angiogenesis mediated diseases
PT   such as solid tumors, comprises recombinantly producing the protein using
PT   an expression system, and recovering and purifying the protein.
XX
PS   Claim 5; Page 30; 67pp; English.
XX
CC   The sequence represents Human Endostatin(TM) C-terminus minus 1 protein,
CC   a natural variant lacking the C-terminal amino acid of Endostatin(TM)
CC   recovered from fermentations of Pichia pastoris cultures harbouring a
CC   expression plasmid containing the Endostatin(TM) DNA sequence given in
CC   AAS00867. The new method of the invention is useful for producing,
CC   recovering and purifying Endostatin (TM) from biological sources, such as
CC   biological fluids, tissues, cells, culture media, and fermentation media.
CC   Endostatin(TM) is useful for treating angiogenesis mediated diseases such
CC   as solid tumours, blood borne tumours, leukaemias, tumour metastases,
CC   benign tumours, e.g. haemangioma, acoustic neuromas, neurofibromas,
CC   trachomas, and pyogenic granulomas, rheumatoid arthritis, psoriasis,
CC   ocular angiogenic diseases, e.g., diabetic retinopathy, retinopathy of
CC   prematurity, macular degeneration, corneal graft rejection, neovascular
CC   glaucoma, colon cancer, retrolental fibroplasia, rubeosis, Osler-Webber
CC   Syndrome, myocardial angiogenesis, plaque neovascularisation,
CC   telangiectasia, haemophiliac joints, angiofibroma, and wound granulation.
CC   Endostatin(TM) is also useful for treating disease of excessive or
CC   abnormal stimulation of endothelial cells such as intestinal adhesions,
CC   atherosclerosis, scleroderma and hypertrophic scars. Higher yields of
CC   more purified, and biologically active Endostatin(TM) are obtained by the
CC   new method. Endostatin(TM) can be stored in buffers for extended periods
CC   of time, and also subjected to lyophilisation, while preserving
CC   biological activity. Centrifugation of broth from fermentation steps in
CC   production is avoided, preventing unwanted potential cellular lysis and
CC   contamination with additional proteins, pigments, enzymes and other
CC   cellular chemicals and debris
XX
SQ   Sequence 182 AA;

     Query Match             86.4%;  Score 835;  DB 4;  Length 182;
     Best Local Similarity   85.1%;  Pred. No. 2.7e-91;
     Matches  154; Conservative   14; Mismatches  13; Indels   0; Gaps   0;

Qy          1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
              |:|:||  ||||||| ||| ||||||||||||| ||||| ||||||||||||||||||||
Db          1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy         61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
              |||||  |||| :||:|:||||||||||||||||||||||||||| :||:||  || |||
Db         61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy        121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
              |||| ||||||.|||||||||:|:||||||||| ||||||  |:  ||.||:|||||:|:
Db        121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy        181 S 181
              :
Db        181 A 181


RESULT 10
AAU77951
ID   AAU77951 standard; protein; 182 AA.
XX
AC   AAU77951;
XX
DT   02-JUL-2002 (first entry)
XX
DE   Amino acid sequence for human endostatin.
XX
KW   Human; immunoconjugate; anti-vascular endothelial growth factor antibody;
KW   anti-VEGF antibody; monoclonal antibody 2C3 ATCC PTA 1595; VEGF receptor;
KW   VEGFR2; KDR/Flk-1; VEGFRi; Flt-1; angiogenesis; macular degeneration;
KW   ocular neovascular disease; cancer; vascularised solid tumour; AIDS;
KW   metastatic tumour; endothelial cell proliferation; inflammatory disorder;
KW   atherosclerosis; diabetic retinopathy; corneal graft rejection;
KW   acquired immune deficiency syndrome; infection; restenosis; fungal ulcer;
KW   sickle cell anaemia; endometriosis; endostatin.
XX
OS   Homo sapiens.
XX
PN   AU200179401-A.
XX
PD   06-DEC-2001.
XX
PF   12-OCT-2001; 2001AU-00079401.
XX
PR   28-APR-2000; 2000AU-00048049.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Thorpe PE,  Brekken RA;
XX
DR   WPI; 2002-281368/33.
XX
PT   Immunoconjugate compositions for treating cancer by inhibiting
PT   angiogenesis and for delivering a diagnostic agent to tumor, comprises
PT   anti-vascular endothelial growth factor antibody attached to a biological
PT   agent.
XX
PS   Example 10; Page 12-13 (Sequence listing); 299pp; English.
XX
CC   The present invention relates to antibody-based compositions comprising
CC   an immunoconjugate such as anti-vascular endothelial growth factor (VEGF)
CC   antibody (Ab) (or its antigen-binding fragment), attached to a biological
CC   agent, where the Ab binds to the same epitope as the monoclonal antibody
CC   (MAb) 2C3 ATCC PTA 1595, and significantly inhibits VEGF binding to the
```

CC   VEGF receptor VEGFR2 (KDR/Flk-1) without inhibiting VEGF binding to the
CC   VEGF receptor VEGFR1 (Flt-1). The compositions of the invention are
CC   useful in therapy, and diagnosis, for inhibiting angiogenesis in an
CC   animal having ocular neovascular disease or macular degeneration, and for
CC   delivering a biological agent to a vascularised tumour. The compositions
CC   can also be used for treating cancer and subjects at risk of developing,
CC   a vascularised solid tumour, a metastatic tumour or metastases from a
CC   primary tumour. The composition is useful for specifically inhibiting
CC   VEGF-induced endothelial cell proliferation, without significantly
CC   inhibiting VEGF-induced macrophage, osteoclast or chondroclast function.
CC   The compositions can be used for treating various diseases such as
CC   inflammatory disorders, atherosclerosis, diabetic retinopathy,
CC   restenosis, acquired immune deficiency syndrome (AIDS), blood borne
CC   tumours, corneal graft rejection, Crohn's disease, fungal ulcers,
CC   infections, sickle cell anaemia, and endometriosis. The present sequence
CC   represents human endostatin. Endostatin may be attached or functionally
CC   associated with anti-VEGF antibodies
XX
SQ   Sequence 182 AA;

Query Match   86.4%; Score 835; DB 5; Length 182;
Best Local Similarity   85.1%; Pred. No. 2.7e-91;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
         |:|:||| ||||||||||| |||||||||||||||||| ||||||||||||||||||||
Db   1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
          |||||||| ||||:|:||||||||||||||:|||||||||||||||||:|| || |||
Db   61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy 121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
          |||||:|||||:|||||||||||| |||||||||| |||:|| |||:||:||||||:|
Db 121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy 181 S 181
          :
Db 181 A 181

RESULT 11
AAY02113
ID   AAY02113 standard; protein; 183 AA.
XX
AC   AAY02113;
XX
DT   16-JUL-1999 (first entry)
XX
DE   SEQ ID 76 of WO9916889.
XX
KW   Angiostatin; endostatin; interferon; thrombospondin;
KW   interferon-inducible protein; platelet factor 4; anti-angiogenic;
KW   anti-tumor; multifunctional protein; angiogenic-mediated disease; cancer;
KW   diabetic retinopathy; macular degeneration; arthritis;
KW   tumor cell production.
XX
OS   Homo sapiens.
XX
PN   WO9916889-A1.
XX
PD   08-APR-1999.
XX
PF   30-SEP-1998; 98WO-US020464.
XX
PR   01-OCT-1997; 97US-0060609P.
XX
PA   (SEAR) SEARLE & CO G D.
XX
PI   Bolanowski MA, Caparon MH, Casperson GF, Gregory SA, Klein BK;
PI   McKearn JP;
XX
DR   WPI; 1999-255098/21.

XX
PT   New multifunctional proteins useful for treating angiogenic-mediated
PT   diseases.
XX
PS   Disclosure; Page 106-107; 121pp; English.
XX
CC   The specification describes multifunctional proteins which comprise
CC   combinations of angiostatin, endostatin, interferon, thrombospondin,
CC   interferon-inducible protein and platelet factor 4, and have anti-
CC   angiogenic and/or anti-tumor activity. The multifunctional protein may
CC   exhibit useful properties such as having similar or greater biological
CC   activity when compared to a single factor or by having improved half-life
CC   or decreased adverse side effects, or a combination of these properties.
CC   The proteins can be used for treating an angiogenic-mediated disease.
CC   e.g. cancer, diabetic retinopathy, macular degeneration, or arthritis.
CC   They can also be used for inhibiting the production of tumor cells
CC   (characteristic of lung, breast, ovarian, prostate, pancreatic, gastric,
CC   colon, renal, bladder cancers; melanoma, hepatoma, sarcoma and lymphoma)
CC   in a patient and for inhibiting tumor growth. The present sequence is
CC   used in the course of the invention
XX
SQ   Sequence 183 AA;

Query Match   86.4%; Score 835; DB 2; Length 183;
Best Local Similarity   85.1%; Pred. No. 2.8e-91;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
         |:|:||| ||||||||||| |||||||||||||||||| |||||||||||||||||||||
Db   1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
          |||||||| ||||:|:||||||||||||||:|||||||||||||||||:|| || |||
Db   61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy 121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
          |||||:|||||:|||||||||||| |||||||||| |||:|| |||:||:||||||:|
Db 121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy 181 S 181
          :
Db 181 A 181

RESULT 12
AAY08693
ID   AAY08693 standard; protein; 183 AA.
XX
AC   AAY08693;
XX
DT   10-AUG-1999 (first entry)
XX
DE   Human endostatin protein fragment.
XX
KW   Plasminogen; human; angiostatin; endostatin; gene therapy; vector;
KW   anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW   tumour growth; solid tumour; diabetic retinopathy; retina.
XX
OS   Homo sapiens.
XX
PN   WO9926480-A1.
XX
PD   03-JUN-1999.
XX
PF   20-NOV-1998; 98WO-US024950.
XX
PR   20-NOV-1997; 97US-0097424.
XX
PA   (GENE-) GENETIX PHARM INC.
PA   (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
PI   Leboulch P, Pawliuk RJ, Bachelot T;
XX

```
DR   WPI; 1999-357696/30.
DR   N-PSDB; AAX77719.
XX
PT   Anti-angiogenic gene therapy vectors.
XX
PS   Disclosure; Page 74-75; 83pp; English.
XX
CC   This invention describes a novel viral gene therapy vector comprising a
CC   nucleic acid molecule encoding an anti-angiogenic polypeptide chosen from
CC   human or murine angiostatin, human or murine endostatin and angiogenesis-
CC   inhibiting fusions and fragments, where the viral vector is sufficiently
CC   attenuated for use in human gene therapy. The products of the invention
CC   have anti-angiogenic, cytostatic, anti-diabetic and ophthalmological
CC   activity. The vector is used in gene therapy for inhibiting tumour growth
CC   in humans harbouring a solid tumour. The vector expresses an anti-
CC   angiogenic polypeptide. An additional use comprises treatment of diabetic
CC   retinopathy, where the anti-angiogenic polypeptide inhibits angiogenesis
CC   in the vicinity of the retina. The vector is administered to cells ex
CC   vivo and then administered to the patient
XX
SQ   Sequence 183 AA;

Query Match       86.4%;    Score 835;  DB 2;  Length 183;
Best Local Similarity   85.1%;   Pred. No. 2.8e-91;
Matches  154;  Conservative  14;  Mismatches  13;  Indels  0;  Gaps  0;

Qy    1  HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI  60
            ::|||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db    1  HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI  60

Qy   61  VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW  120
            ||||||:|||||||||||||||||||||||||||||||||||||:||  :::||||
Db   61  VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW  120

Qy  121  HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT  180
           ||||::|||||:||||||||||:|||||||||||||||:||  |:::||||||||||| ||
Db  121  HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT  180

Qy  181  S  181
          :
Db  181  A  181

RESULT 13
AAY70252
ID   AAY70252 standard; protein; 183 AA.
XX
AC   AAY70252;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Human angiogenesis inhibitor, endostatin.
XX
KW   Human; immunoglobulin gamma Fc fragment; endostatin; immunofusin;
KW   angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;
KW   antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;
KW   vasotropic; vulnerary; treatment; antiarteriosclerosis; tumour;
KW   metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
KW   ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW   myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW   wound granulation; keloid scar; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO20001103-A2.
XX
PD   02-MAR-2000.
XX
PF   25-AUG-1999;  99WO-US019329.
XX
PR   25-AUG-1998;  98US-0097883P.
XX
PA   (LEXI-) LEXINGEN PHARM CORP.
```

```
XX
PI   Lo K, Li Y, Gillies SD;
XX
DR   WPI; 2000-237616/20.
DR   N-PSDB; AAZ51291.
XX
PT   Novel fusion protein of angiostatin or endostatin and an immunoglobulin
PT   FC region, useful for treating conditions mediated by angiogenesis, such
PT   as rheumatoid arthritis, tumors and macular degeneration.
PS   Example 1; Page 41-42; 68pp; English.
XX
CC   The patent discloses a DNA molecule encoding a fusion protein comprising
CC   a signal sequence, an immunoglobulin Fc region, and an angiogenesis
CC   inhibitor selected from angiostatin, endostatin, a plasminogen fragment
CC   having angiostatin activity, a collagen XVIII fragment having endostatin
CC   activity, or combinations of them. The fusion protein (immunofusin) is
CC   used to inhibit angiogenesis and to treat diseases or conditions mediated
CC   by angiogenesis. Conditions that may be treated include solid tumours,
CC   blood born tumours, tumour metastasis, benign tumours including
CC   haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic
CC   granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
CC   e.g. diabetic retinopathy, retinopathy of prematurity, macular
CC   degeneration, corneal graft rejection, neovascular glaucoma, retrolental
CC   fibroplasia, rubeosis and Osler-Webber syndrome; myocardial angiogenesis,
CC   plaque neovascularisation, telangiectasia, haemophiliac joints,
CC   angiofibroma, wound granulation, and excessive or abnormal stimulation of
CC   endothelial cells, intestinal cells, atherosclerosis, sclerodermal and
CC   hypertrophic scars, i.e. keloid scars. The DNA constructs may be used in
CC   gene therapy. The present sequence is a human endostatin used in the
CC   construction of immunofusin containing human immunoglobulin gamma (IgG)
CC   Fc fragment
XX
SQ   Sequence 183 AA;

Query Match       86.4%;    Score 835;  DB 3;  Length 183;
Best Local Similarity   85.1%;   Pred. No. 2.8e-91;
Matches  154;  Conservative  14;  Mismatches  13;  Indels  0;  Gaps  0;

Qy    1  HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI  60
            ::|||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db    1  HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI  60

Qy   61  VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW  120
            ||||||:|||||||||||||||||||||||||||||||||||||:||  :::||||
Db   61  VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW  120

Qy  121  HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT  180
           ||||::|||||:||||||||||:|||||||||||||||:||  |:::||||||||||| ||
Db  121  HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT  180

Qy  181  S  181
          :
Db  181  A  181

RESULT 14
AAY90771
ID   AAY90771 standard; protein; 183 AA.
XX
AC   AAY90771;
XX
DT   22-AUG-2000  (first entry)
XX
DE   Human angiogenesis inhibiting factor 1 protein.
XX
KW   Human; angiogenesis inhibiting factor 1; IAF-1; tumour; antibody;
KW   abnormal vessel disease.
XX
OS   Homo sapiens.
XX
PN   CN1244536-A.
XX
```

PD   16-FEB-2000.
XX
PF   10-AUG-1998;   98CN-00117150.
XX
PR   10-AUG-1998;   98CN-00117150.
XX
PA   (ONCO-) INST ONCOLOGY UNDER TUMOR HOSPITAL CHINE.
XX
PI   Yang Z, Guo W;
XX
DR   WPI; 2000-388168/34.
DR   N-PSDB; AAA29884.
XX
PT   Angiogenesis inhibiting factor 1 and its derivative useful for treating
PT   tumors.
XX
PS   Claim 1; Fig 5; 41pp; Chinese.
XX
CC   The present sequence represents an angiogenesis inhibiting factor (I),
CC   designated IAF-1. The present invention also describes: (1) preparation
CC   of (I) and its derivative; (2) an IAF binding acceptor and its
CC   preparation; and (3) an IAF antibody. (I) is useful for preparing new
CC   biological preparations for effectively treating various tumours and
CC   abnormal-vessel diseases. The IAF antibody is preferably a polyclonal
CC   antibody, mosaic antibody, single stranded antibody and human originated
CC   antibody
XX
SQ   Sequence 183 AA;

Query Match   86.4%; Score 835; DB 3; Length 183;
Best Local Similarity   85.1%; Pred. No. 2.8e-91;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
             ::||||||||||||||:||||||||||||||||:||||||||||||||||
Db   1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
             ||||| ||:||:|||:||||||||||||||||||||||||:|:||||
Db   61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy   121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
             ||||:|||||:|||||||||| :||||||||||::||||:::||||||||||
Db   121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy   181 S 181
             :
Db   181 A 181

RESULT 15
AAB16451
ID   AAB16451 standard; protein; 183 AA.
XX
AC   AAB16451;
XX
DT   27-OCT-2000 (first entry)
XX
DE   Human endostatin protein sequence.
XX
KW   Angiogenesis-inhibiting protein receptor; angiogenesis; angiostatin;
KW   endostatin; plasminogen; laminin; treatment; wound healing; solid tumour;
KW   psoriasis; scleroderma; myocardial angiogenesis; Crohn's disease;
KW   cerebral collateral; arteriovenous malformation; rubeosis; cancer;
KW   diabetic retinopathy; arthritis; wound healing; peptic ulcer;
KW   Helicobacter related disease; fracture; cat scratch fever.
XX
OS   Homo sapiens.
XX
PN   WO200032631-A2.
XX
PD   08-JUN-2000.
XX

PF   06-DEC-1999;   99WO-US028897.
XX
PR   04-DEC-1998;   98US-0206059.
XX
PA   (ENTR-) ENTREMED INC.
XX
PI   Macdonald NJ, Sim KL;
XX
DR   WPI; 2000-412290/35.
XX
PT   New angiogenesis-inhibiting protein receptors, and useful in methods for
PT   treating diseases and processes that are mediated by angiogenesis, such
PT   as solid tumors, psoriasis, scleroderma and myocardial angiogenesis.
XX
PS   Disclosure; Fig 3; 100pp; English.
XX
CC   This invention relates to angiogenesis-inhibiting protein receptors, and
CC   the DNA sequences encoding them. Angiogenesis is the generation of new
CC   blood vessels into a tissue, and normally occurs in wound healing, foetal
CC   and embryonal development and the formation of the corpus luteum,
CC   endometrium and placenta. Angiostatin is a protein (see AAB16450 and
CC   AAA68202) involved in angiogenesis, and has an amino acid sequence
CC   similar to that of a plasminogen fragment (see murine plasminogen
CC   AAB16490). Angiostatin has the ability to inhibit angiogenesis.
CC   Endostatin is also an angiogenesis inhibiting protein (see AAB16451 and
CC   AAA68203). Sequences AAA68242 and AAB16522 represent coding and protein
CC   sequences of human laminin. Laminin is an angiostatin binding protein,
CC   and some of the peptides of the invention share homology with regions of
CC   laminin. Peptides AAB16452-B16521 (excluding AAB16490) are the
CC   angiogenesis-inhibiting protein receptor fragments of the invention. The
CC   peptides bind either angiostatin or endostatin and can be used in methods
CC   for treating diseases and processes that are mediated by angiogenesis,
CC   such as solid tumours, psoriasis, scleroderma, myocardial angiogenesis,
CC   Crohn's disease, cerebral collaterals, arteriovenous malformations,
CC   rubeosis, diabetic retinopathy, arthritis, wound healing, peptic ulcers,
CC   Helicobacter related diseases, fractures, placentation and cat scratch
CC   fever. They are useful for the detection and prognosis of cancer. DNA
CC   sequences A628204-A628241 encode the peptides of the invention
XX
SQ   Sequence 183 AA;

Query Match   86.4%; Score 835; DB 3; Length 183;
Best Local Similarity   85.1%; Pred. No. 2.8e-91;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
             ::||||||||||||||:||||||||||||||||:||||||||||||||||
Db   1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
             ||||| ||:||:|||:||||||||||||||||||||||||:|:||||
Db   61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy   121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
             ||||:|||||:|||||||||| :||||||||||::||||:::||||||||||
Db   121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy   181 S 181
             :
Db   181 A 181

Search completed: March 26, 2004, 13:35:58
Job time : 47.7778 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      March 26, 2004, 13:34:07 ; Search time 14.6667 Seconds
                                         (without alignments)
                                         647.671 Million cell updates/sec

Title:       US-09-938-391-4
Perfect score: 966
Sequence:    1 HTHQDFQLVLHLVALNSPQP.........CRHAFVVLCIENSVMTSFSK 184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:    389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:       389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued Patents AA:*
             1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
             2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
             3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
             4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
             5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
             6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	835	86.4	182	4	US-09-561-500-14	Sequence 14, Appl
2	835	86.4	182	4	US-09-561-108-14	Sequence 14, Appl
3	835	86.4	182	4	US-09-315-689-3	Sequence 3, Appl
4	835	86.4	182	4	US-09-561-526-14	Sequence 14, Appl
5	835	86.4	182	4	US-09-561-499-14	Sequence 14, Appl
6	835	86.4	182	4	US-09-998-831-14	Sequence 14, Appl
7	835	86.4	183	3	US-09-206-059-2	Sequence 2, Appl
8	831	86.0	191	4	US-09-561-500-13	Sequence 13, Appl
9	831	86.0	191	4	US-09-561-108-13	Sequence 13, Appl
10	831	86.0	191	4	US-09-561-526-13	Sequence 13, Appl
11	831	86.0	191	4	US-09-561-499-13	Sequence 13, Appl
12	831	86.0	191	4	US-09-998-831-13	Sequence 13, Appl
13	828	85.7	195	1	US-08-159-784-2	Sequence 2, Appli
14	817	84.6	178	5	US-09-315-689-5	Sequence 5, Appl
15	790	81.8	185	3	US-08-985-526-36	Sequence 36, Appl
16	513	53.1	191	4	US-09-998-831-3	Sequence 3, Appli
17	195.5	20.2	124	4	US-09-231-077D-10	Sequence 10, Appl
18	188	19.5	123	4	US-09-231-077D-11	Sequence 11, Appl
19	148	15.3	35	3	US-09-046-985-2	Sequence 2, Appli
20	148	15.3	35	3	US-09-474-743-2	Sequence 2, Appli
21	101	10.5	22	3	US-09-046-985-7	Sequence 7, Appli
22	101	10.5	22	3	US-09-474-743-7	Sequence 7, Appli
23	97	10.0	16	3	US-09-385-442-32	Sequence 32, Appl
24	87	9.0	512	4	US-09-252-991A-17586	Sequence 17586, A
25	86	8.9	20	3	US-08-740-168A-1	Sequence 1, Appli
26	86	8.9	20	2	US-09-349-429-1	Sequence 1, Appli
27	86	8.9	20	4	US-09-315-689-1	Sequence 1, Appli
28	86	8.9	20	4	US-09-174-282-1	Sequence 1, Appli
29	86	8.9	20	4	US-09-154-302-1	Sequence 1, Appli
30	83.5	8.6	403	4	US-09-252-991A-17378	Sequence 17378, A
31	83	8.6	682	4	US-09-489-039A-10595	Sequence 10595, A
32	82	8.5	428	4	US-09-252-991A-23863	Sequence 23863, A
33	81.5	8.4	182	4	US-09-252-991A-32078	Sequence 32078, A
34	79.5	8.2	690	4	US-09-252-991A-32350	Sequence 32350, A
35	75.5	7.8	336	3	US-09-105-537-27	Sequence 27, Appl
36	75.5	7.8	11877	3	US-09-105-537-6	Sequence 6, Appli
37	75	7.8	347	4	US-09-252-991A-25606	Sequence 25606, A
38	75	7.8	419	4	US-09-252-991A-30853	Sequence 30853, A
39	75	7.8	654	4	US-09-907-794A-177	Sequence 177, App
40	75	7.8	654	4	US-09-905-125A-177	Sequence 177, App
41	75	7.8	654	4	US-09-902-775A-177	Sequence 177, App
42	75	7.8	1064	4	US-09-252-991A-17508	Sequence 17508, A
43	74.5	7.7	298	4	US-09-252-991A-26179	Sequence 26179, A
44	74.5	7.7	360	4	US-09-252-991A-26045	Sequence 26045, A
45	74.5	7.7	410	4	US-09-252-991A-31937	Sequence 31937, A

ALIGNMENTS

```
RESULT 1
US-09-561-500-14
; Sequence 14, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
;  APPLICANT: Philip E. Thorpe
;  APPLICANT: Rolf A. Brekken
;  TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
;  FILE REFERENCE: 4001.002500
;  CURRENT APPLICATION NUMBER: US/09/561,500
;  CURRENT FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/131,432
;  PRIOR FILING DATE: 1999-04-28
;  NUMBER OF SEQ ID NOS:   44
;  SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 14
;  LENGTH: 182
;  TYPE: PRT
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
;  OTHER INFORMATION: PEPTIDE
US-09-561-500-14

Query Match             86.4%;  Score 835;  DB 4;  Length 182;
Best Local Similarity   85.1%;  Pred. No. 7e-96;
Matches  154;  Conservative  14;  Mismatches  13;  Indels   0;  Gaps   0;

Qy      1 HTHQDFQLVLHLVALNSPQPGPGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
            |:|:||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy     61 VRRADRTGVPVVNLRDEVLFPSWEALFPSGSEGQLKPGARIFSFDGKRDVLQHPAWPRKSVW 120
           |:||||::||||||:|||||||||||||||||||:|::|||||||||||||:|:|::|||:|
Db     61 VRRADRAAVPIVNLKDELLFPSWEALFPSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy    121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRILLEQEAASCRHAFVVLCIENSVMT 180
           ||||:||||||:::|||||||||||||||||||||||||||||::||||||||||||||||
Db    121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRILLGQSAASCHHAYIVLCIENSFMT 180

Qy    181 S 181
          :
Db    181 A 181

RESULT 2
US-09-561-108-14
; Sequence 14, Application US/09561108
; Patent No. 6342221
; GENERAL INFORMATION:
```

; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/561,108
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-561-108-14

Query Match   86.4%;   Score 835; DB 4; Length 182;
Best Local Similarity   85.1%;   Pred. No. 7e-96;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
            |:|:||| ||||||||||||| |||||||||||||||||||||||||||||||||||
Db   1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
            ||||| |:||||||||||:||||||||||||:||||||||||| |:||:| ||
Db   61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy   121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
            ||||||||||||||||||||||||||||||||:||||||| ||:||||||| |
Db   121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy   181 S 181
             :
Db   181 A 181

RESULT 3
US-09-315-689-3
; Sequence 3, Application US/09315689
; Patent No. 6346510
; GENERAL INFORMATION:
; APPLICANT: Folkman, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
; FILE REFERENCE: 05213-0229
; CURRENT APPLICATION NUMBER: US/09/315,689
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-315-689-3

Query Match   86.4%;   Score 835; DB 4; Length 182;
Best Local Similarity   85.1%;   Pred. No. 7e-96;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
            |:|:||| ||||||||||||| |||||||||||||||||||||||||||||||||||
Db   1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
            ||||| |:||||||||||:||||||||||||:||||||||||| |:||:| ||
Db   61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy   121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
            ||||| ||||| |||| ||||||||| ||||||| |||| | ||:||||||| |
Db   121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGRLLGQSAASCHHAYIVLCIENSFMT 180

Db   121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy   181 S 181
             :
Db   181 A 181

RESULT 4
US-09-561-526-14
; Sequence 14, Application US/09561526
; Patent No. 6416758
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002586
; CURRENT APPLICATION NUMBER: US/09/561,526
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-561-526-14

Query Match   86.4%;   Score 835; DB 4; Length 182;
Best Local Similarity   85.1%;   Pred. No. 7e-96;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
            |:|:||| ||||||||||||| |||||||||||||||||||||||||||||||||||
Db   1 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
            ||||| |:||||||||||:||||||||||||:||||||||||| |:||:| ||
Db   61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy   121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
            ||||||||||| |||| ||||||||| ||||||| |:|||||| ||:||||||| |
Db   121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy   181 S 181
             :
Db   181 A 181

RESULT 5
US-09-561-499-14
; Sequence 14, Application US/09561499
; Patent No. 6524583
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/09/561,499
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

; OTHER INFORMATION: PEPTIDE
US-09-561-499-14

```
Query Match           86.4%;   Score 835;  DB 4;  Length 182;
Best Local Similarity 85.1%;   Pred. No. 7e-96;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGWRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
         |:||:|||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db   1 HSHRDFQPVLHLVALNSPLSGGWRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy  61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
         |||||| |||||:|:||:||||||||||||||||||||||||||||||:||:||
Db  61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy 121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
         ||||||||||:|||||||||||||:|||||||||||||||||||:||||||||:|||
Db 121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy 181 S 181
         :
Db 181 A 181
```

RESULT 6
US-09-998-831-14
; Sequence 14, Application US/09998831
; Patent No. 6676941
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
; TITLE OF INVENTION: INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/998,831
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 09/561,108
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-998-831-14

```
Query Match           86.4%;   Score 835;  DB 4;  Length 182;
Best Local Similarity 85.1%;   Pred. No. 7e-96;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGWRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
         |:||:|||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db   1 HSHRDFQPVLHLVALNSPLSGGWRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy  61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
         |||||| |||||:|:||:||||||||||||||||||||||||||||||:||:||
Db  61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy 121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
         ||||||||||:|||||||||||||:|||||||||||||||||||:||||||||:|||
Db 121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy 181 S 181
         :
Db 181 A 181
```

RESULT 7
US-09-206-059-2
; Sequence 2, Application US/09206059
; Patent No. 6201104
; GENERAL INFORMATION:
; APPLICANT: MacDonald, Nicholas
; APPLICANT: Sim, Kim Lee
; TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and
; TITLE OF INVENTION: Proteins and Methods of Use
; FILE REFERENCE: 05213-0370
; CURRENT APPLICATION NUMBER: US/09/206,059
; CURRENT FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-206-059-2

```
Query Match           86.4%;   Score 835;  DB 3;  Length 183;
Best Local Similarity 85.1%;   Pred. No. 7.1e-96;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGWRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
         |:||:|||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db   1 HSHRDFQPVLHLVALNSPLSGGWRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 60

Qy  61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
         |||||| |||||:|:||:||||||||||||||||||||||||||||||:||:||
Db  61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 120

Qy 121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
         ||||||||||:|||||||||||||:|||||||||||||||||||:||||||||:|||
Db 121 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180

Qy 181 S 181
         :
Db 181 A 181
```

RESULT 8
US-09-561-500-13
; Sequence 13, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002500
; CURRENT APPLICATION NUMBER: US/09/561,500
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-500-13

```
Query Match           86.0%;   Score 831;  DB 4;  Length 191;
Best Local Similarity 83.7%;   Pred. No. 2.4e-95;
Matches 154; Conservative 17; Mismatches 13; Indels 0; Gaps 0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGWRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
         |:||:|||||||||||||||:|||||||||||||||||:|||||||||||||||||||||
Db   8 HTHQDFQPVLHLVALNTPLSGGWRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSI 67

Qy  61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
         ||||:| |||:||:|||:|||||||||||:|:|||||||||||||||||:||:|||:||
Db  68 VRRADRGSVPIVNLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQKSVW 127
```

```
Qy   121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
             ||||||||||||||:|||||||||||:|||| :||:|||||| :::|||||||||
Db   128 HGSDPSGRRLMESYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMT 187

Qy   181 SFSK 184
             ||||
Db   188 SFSK 191

RESULT 9
US-09-561-108-13
; Sequence 13, Application US/09561108
; Patent No. 6342221
; GENERAL INFORMATION:
;   APPLICANT:  Philip E. Thorpe
;   APPLICANT:  Rolf A. Brekken
;   TITLE OF INVENTION:  ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
;   FILE REFERENCE:  4001.002584
;   CURRENT APPLICATION NUMBER: US/09/561,108
;   CURRENT FILING DATE:  2000-04-28
;   PRIOR APPLICATION NUMBER:  60/131,432
;   PRIOR FILING DATE:  1999-04-28
;   NUMBER OF SEQ ID NOS:  44
;   SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 191
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
US-09-561-108-13

Query Match      86.0%;   Score 831;  DB 4;  Length 191;
Best Local Similarity  83.7%;   Pred. No. 2.4e-95;
Matches  154;  Conservative  17;  Mismatches  13;  Indels  0;  Gaps  0;

Qy     1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
           ||||||:|||||||:|||| :|||||||||||||:|||||||||||:||||||||||||
Db     8 HTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSI 67

Qy    61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
           |||||||:||||:|||||||||:||:|||||:|||| :|||||||||| :||||:||:|:|
Db    68 VRRADRGSVPIVNLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQKSVW 127

Qy   121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
           ||||||||||||:|||||||||||:|||| :||:|||||| :::|||||||||
Db   128 HGSDPSGRRLMESYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMT 187

Qy   181 SFSK 184
           ||||
Db   188 SFSK 191

RESULT 10
US-09-561-526-13
; Sequence 13, Application US/09561526
; Patent No. 6416758
; GENERAL INFORMATION:
;   APPLICANT:  Philip E. Thorpe
;   APPLICANT:  Rolf A. Brekken
;   TITLE OF INVENTION:  ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
;   FILE REFERENCE:  4001.002586
;   CURRENT APPLICATION NUMBER: US/09/561,526
;   CURRENT FILING DATE:  2000-04-28
;   PRIOR APPLICATION NUMBER:  60/131,432
;   PRIOR FILING DATE:  1999-04-28
;   NUMBER OF SEQ ID NOS:  44
;   SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 191
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
```

```
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
US-09-561-526-13

Query Match      86.0%;   Score 831;  DB 4;  Length 191;
Best Local Similarity  83.7%;   Pred. No. 2.4e-95;
Matches  154;  Conservative  17;  Mismatches  13;  Indels  0;  Gaps  0;

Qy     1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
           ||||||:|||||||:|||| :|||||||||||||:|||||||||||:||||||||||||
Db     8 HTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSI 67

Qy    61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
           |||||||:||||:|||||||||:||:|||||:|||| :|||||||||| :||||:||:|:|
Db    68 VRRADRGSVPIVNLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQKSVW 127

Qy   121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
           ||||||||||||:|||||||||||:|||| :||:|||||| :::|||||||||
Db   128 HGSDPSGRRLMESYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMT 187

Qy   181 SFSK 184
           ||||
Db   188 SFSK 191

RESULT 11
US-09-561-499-13
; Sequence 13, Application US/09561499
; Patent No. 6524583
; GENERAL INFORMATION:
;   APPLICANT:  Philip E. Thorpe
;   APPLICANT:  Rolf A. Brekken
;   TITLE OF INVENTION:  ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
;   FILE REFERENCE:  4001.002582
;   CURRENT APPLICATION NUMBER: US/09/561,499
;   CURRENT FILING DATE:  2000-04-28
;   NUMBER OF SEQ ID NOS:  44
;   PRIOR APPLICATION NUMBER:  60/131,432
;   PRIOR FILING DATE:  1999-04-28
;   SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 191
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
US-09-561-499-13

Query Match      86.0%;   Score 831;  DB 4;  Length 191;
Best Local Similarity  83.7%;   Pred. No. 2.4e-95;
Matches  154;  Conservative  17;  Mismatches  13;  Indels  0;  Gaps  0;

Qy     1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
           ||||||:|||||||:|||| :|||||||||||||:|||||||||||:||||||||||||
Db     8 HTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSI 67

Qy    61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
           |||||||:||||:|||||||||:||:|||||:|||| :|||||||||| :||||:||:|:|
Db    68 VRRADRGSVPIVNLKDEVLSPSWDSLFSGSQGQLQPGARIFSFDGRDVLRHPAWPQKSVW 127

Qy   121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
           ||||||||||||:|||||||||||:|||| :||:|||||| :::|||||||||
Db   128 HGSDPSGRRLMESYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMT 187

Qy   181 SFSK 184
           ||||
Db   188 SFSK 191

RESULT 12
US-09-998-831-13
; Sequence 13, Application US/09998831
; Patent No. 6676941
; GENERAL INFORMATION:
```

APPLICANT: Philip E. Thorpe
APPLICANT: Rolf A. Brekken
TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
TITLE OF INVENTION: INHIBITING VEGF
FILE REFERENCE: 4001.002584
CURRENT APPLICATION NUMBER: US/09/998,831
CURRENT FILING DATE: 2001-11-30
PRIOR APPLICATION NUMBER: 09/561,108
PRIOR FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 191
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-998-831-13

Query Match   86.0%;   Score 831; DB 4; Length 191;
Best Local Similarity 83.7%;   Pred. No. 2.4e-95;
Matches 154; Conservative 17; Mismatches 13; Indels 0; Gaps 0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGMWRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
          ||||||||||||||||||||||:||: ||||||||||||| |:|||||||||||||||||
Db   8 HTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSI 67

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
          |||||| |:||:|| |||:::|||:||||||:||||||||||:||:||| ||||||
Db   68 VRRADRGSVPIVNLKDEVLSPSWDSLFSGSQQLQPGARIFSFDGRDVLRHPAWPQKSVW 127

Qy 121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQRAASCRHAFVVLCIENSVMT 180
          ||||||||||:|||||||||:||||||||||||| |||||||||::||| ::|||||| |
Db 128 HGSDPSGRRLMESYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMT 187

Qy 181 SFSK 184
          ||||
Db 188 SFSK 191

RESULT 13
US-08-159-784-2
; Sequence 2, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070

; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 195
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
US-08-159-784-2

Query Match   85.7%;   Score 828; DB 1; Length 195;
Best Local Similarity 83.2%;   Pred. No. 5.8e-95;
Matches 153; Conservative 18; Mismatches 13; Indels 0; Gaps 0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGMWRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
          ||||||||||||||||||||||:||: ||||||||||||| |:|||||||||||||||||
Db   12 HTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSI 71

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
          |||||| |:||:|| |||:::|||:||||||:||||||||||:||:||| ||||||
Db   72 VRRADRGSVPIVNLKDEVLSPSWDSLFSGSQQVQPGARIFSFDGRDVLRHPAWPQKSVW 131

Qy 121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQRAASCRHAFVVLCIENSVMT 180
          ||||||||||:|||||||||:||||||||||||| |||||||||::||| ::|||||| |
Db 132 HGSDPSGRRLMESYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMT 191

Qy 181 SFSK 184
          ||||
Db 192 SFSK 195

RESULT 14
US-09-315-689-5
; Sequence 5, Application US/09315689
; Patent No. 6345510
; GENERAL INFORMATION:
; APPLICANT: Folkman, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
; FILE REFERENCE: 05213-0229
; CURRENT APPLICATION NUMBER: US/09/315,689
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-315-689-5

Query Match   84.6%;   Score 817; DB 4; Length 178;
Best Local Similarity 85.9%;   Pred. No. 1.2e-93;
Matches 152; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

Qy   5 DFQLVLHLVALNSPQPGGMWRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRA 64
          ||||||||||||||||||||:||: ||||||||||||| |:||||||||||||||||||||
Db   1 DFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSIVRRA 60

Qy   65 DRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSD 124
          | |||||||:|:|| |||:::|||:||||||:|||||||||:||:||: ||||||||||
Db   61 DRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVWHGSD 120

Qy 125 PSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQRAASCRHAFVVLCIENSVMTS 181
          ||||||:|||||||||:||||||||||||| |||||||||::||||::|||||||||
Db 121 PNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGGRLLGQSAASCHHAYIVLCIENSFMTA 177

RESULT 15
US-08-985-526-36
; Sequence 36, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A

TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
TITLE OF INVENTION: THERAPY
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
    ADDRESSEE: Connolly, Bove, Lodge, & Hutz
    STREET: 1220 Market Street, P.O. Box 2207
    CITY: Wilmington
    STATE: Delaware
    COUNTRY: U.S.A.
    ZIP: 19899
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/985,526
    FILING DATE:
    CLASSIFICATION:
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/608,845
    FILING DATE: 16-JUL-1996
ATTORNEY/AGENT INFORMATION:
    NAME: McMorrow Jr., Robert G
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (302) 658-9141
    TELEFAX: (302) 658-5613
INFORMATION FOR SEQ ID NO: 36:
    SEQUENCE CHARACTERISTICS:
    LENGTH: 185 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
US-08-985-526-36

Query Match   81.8%;   Score 790; DB 3;   Length 185;
Best Local Similarity   81.1%; Pred. No. 2.9e-90;
Matches 150; Conservative   16; Mismatches   17; Indels   2; Gaps   2;

Qy   1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
           |||||||:||||||||| :||||||||||||||| || :| | |||||||||||||||||
Db   2 HTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCFNNAR-VGLSGTFRAFLSSRLQDLYSI 60

Qy   61 VRRADRTGVPVV-NLRDEVLFPSWEALFSGSBEGQLKPGARIFSFDGRDVLQHPAWPRKSV 119
           || ||||: |: :||||| |:||:::||| ||||||:|||||||||||||||| :||: |
Db   61 VRRADRGSVPIVQNLRDEVLSPSWDSLFSGSQGGLQPGARIFSFDGRDVLRHPAWPQRSV 120

Qy   120 WHGSDPSGRRLTDSYCETWRTEAPAATGQASSLIAGRLLEQEAASCRHAFVVLCIENSVM 179
           |||||||||| :||||||||| :| ||||||||||||||| :::|||:| ||:|||||
Db   121 WHGSDPSGRRLMESYCETWRTETTGATGQASLLSGRLLEQRAASCHDSYIVLCIENSFM 180

Qy   180 TSFSK 184
           |||:|
Db   181 TSFSR 185

Search completed: March 26, 2004, 13:39:22
Job time : 14.6667 secs

```
                    GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    March 26, 2004, 13:33:07 ; Search time 12.4444 Seconds
                                (without alignments)
                            1422.260 Million cell updates/sec

Title:         US-09-938-391-4
Perfect score: 966
Sequence:      1 HTHQDFQLVLHLVALNSPQP.........CRHAFVVLCIENSVMTGFSK 184

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:        283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     PIR_78:*
               1:  pir:*
               2:  pir2:*
               3:  pir3:*
               4:  pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match	Length	DB	ID	Description
1	835	86.4		684	2	A53019	collagen alpha 1(X
2	831	86.0		1774	2	B56101	collagen alpha 1(X
3	828	85.7		1315	2	A56101	collagen alpha 1(X
4	526	54.5		1388	2	A53317	collagen alpha 1(X
5	366	37.9		650	2	T22002	hypothetical prote
6	87	9.0		494	2	A83204	alginate biosynthe
7	82	8.5		995	2	H59432	RhoGAP protein hom
8	81.5	8.4		190	2	T36950	probable transcrip
9	81.5	8.4		633	2	G90704	probable Vgr prote
10	81.5	8.4		633	2	B85555	probable Vgr prote
11	81	8.4		312	2	A55461	NAD(P)-arginine AD
12	79.5	8.2		346	2	G83127	conserved hypothet
13	79.5	8.2		713	2	D90658	VgrG protein [impo
14	79.5	8.2		713	2	D85509	hypothetical prote
15	79	8.2		672	2	T36083	hypothetical prote
16	78.5	8.1		244	2	C45729	sulfolipid biosynt
17	77	8.0		314	2	AG3054	succinoglycan bios
18	77	8.0		314	2	F98231	succinoglycan bios
19	77	8.0		419	2	D42725	nitrile hydratase
20	77	8.0		800	2	AG0438	probable Rhs acces
21	76	7.9		404	2	G90781	hypothetical prote
22	76	7.9		404	2	D85642	hypothetical prote
23	76	7.9		636	2	T10569	probable serine/th
24	76	7.9		732	2	AF0439	probable Rhs acces
25	75.5	7.8		297	2	AE3128	carbohydrate kinas
26	75.5	7.8		319	2	D98159	hypothetical prote
27	75.5	7.8		336	2	T17408	rRNA (adenine-N6-)
28	75.5	7.8		385	2	T51127	GGPP synthase [imp
29	75.5	7.8		513	1	A45333	exopolyphosphatase
30		75.5	7.8	513	2	D91049	exopolyphosphatase
31		75.5	7.8	513	2	H85893	exopolyphosphatase
32		74.5	7.7	702	2	D90886	VgrE protein [impo
33		74.5	7.7	714	2	B85732	Rhs element associ
34		74	7.7	126	2	A69177	hypothetical prote
35		74	7.7	636	2	JW0047	class I cytokinase
36		73.5	7.6	437	2	H95857	probable deaminase
37		73.5	7.6	723	2	AG1662	beta-glucosidases
38		73.5	7.6	807	2	F64844	ycdS protein precu
39		73.5	7.6	807	2	F90787	probable outer mem
40		73.5	7.6	807	2	F85647	probable outer mem
41		73	7.6	427	2	A53798	58K membrane-assoc
42		73	7.6	449	2	H83944	glutamine syntheta
43		73	7.6	996	2	G87687	hypothetical prote
44		72.5	7.5	510	2	G87369	tryptophan halogen
45		72	7.5	286	2	E97241	hypothetical prote

ALIGNMENTS

```
RESULT 1
A53019
collagen alpha 1(XVIII) chain - human (fragment)
N;Contains: endostatin
C;Species: Homo sapiens (man)
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 15-Sep-2003
C;Accession: A53019
R;Oh, S.P.; Warman, M.L.; Seldin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olsen,
Genomics 19, 494-499, 1994
A;Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localiz
A;Reference number: A53019; MUID:94245237; PMID:8188291
A;Accession: A53019
A;Molecule type: mRNA
A;Residues: 1-684 <OHA>
A;Cross-references: GB:L22548; NID:g348908; PIDN:AAA51864.1; PID:g562794
A;Note: the cited accession number, L25548, is not in Genbank release 103
A;Note: in the authors' translation, 482-Gly is at the third position of the tripeptide repeating unit
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
lated and subsequently O-glycosylated.
C;Comment: Different splice forms of collagen alpha 1(XVIII) may be involved in perivas
ay be useful in treating solid tumors.
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of u
C;Genetics:
A;Gene: GDB:COL18A1
A;Cross-references: GDB:138752; OMIM:120328
A;Map position: 21q22.3-21q22.3
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteogly
F;1-684/Product: collagen alpha1(XVIII) chain (fragment) #status predicted <MAT>
F;1-59/Domain: collagenous (fragment) #status predicted <CO4>
F;74-115/Domain: collagenous #status predicted <CO5>
F;129-201/Domain: collagenous #status predicted <CO6>
F;212-244/Domain: collagenous #status predicted <CO7>
F;257-278/Domain: collagenous #status predicted <CO8>
F;262-264/Region: cell attachment (R-G-D) motif
F;286-340/Domain: collagenous #status predicted <CO9>
F;352-371/Domain: collagenous #status predicted <CO10>
F;502-684/Product: endostatin #status predicted <EST>
F;509-684/Region: multiplexin collagen carboxyl-terminal homologous

Query Match          86.4%;   Score 835;   DB 2;  Length 684;
Best Local Similarity 85.1%;   Pred. No. 4.9e-74;
Matches  154; Conservative  14; Mismatches  13;  Indels   0;  Gaps   0;

Qy    1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
          |:|:|| ||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db  502 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 561

Qy   61 VRRADRTGVPVNLRDEVLFPSWEALFSGSEGQLKPGARIFSPDGRDVLQHPAWPRKSVW 120
          |||||| |||:|:|||:|:||||||||||||||||||||||||||||||||||| |:||
Db  562 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPQKSVW 621

Qy  121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
```

Db   622 HGSDPNGRRLTESYCETWRTEAPSATCQASSLLGQSAASCHAYIVLCIENSFMT 681

Qy   181 S 181

Db   682 A 682

RESULT 2
B56101
collagen alpha 1(XVIII) chain precursor, long splice form - mouse
N;Contains: collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin
C;Species: Mus musculus (house mouse)
C;Date: 03-Oct-1995 #sequence revision 08-May-1998 #text_change 15-Sep-2003
C;Accession: B56101; C56101; S72450; S65595; PN0675; A54072; A58816
R;Rehn, M.; Pihlajaniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A;Title: Identification of three N-terminal ends of type XVIII collagen chains and tissu
tif homologous to rat and Drosophila frizzled proteins.
A;Reference number: A56101; MUID:95181468; PMID:7876242
A;Accession: B56101
A;Molecule type: mRNA
A;Residues: 1-562 <REH1>
A;Cross-references: GB:U11637; NID:g618429; PIDN:AAC52179.1; PID:g618430
A;Experimental source: splice form clone PE17.24
A;Accession: C56101
A;Molecule type: mRNA
A;Residues: 1-239, 487-562 <REH2>
A;Cross-references: GB:U11637; NID:g618429
A;Experimental source: splice form clones PE8.1, PE19, PE15.2
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A;Reference number: S72450
A;Accession: S72450
A;Molecule type: mRNA
A;Residues: 487-1146,'L',1148-1193,'F',1195-1210,'R',1212-1512,'L',1514-1522,'F',1524-16
A;Cross-references: EMBL:L22545; NID:g348968; PIDN:AAA19787.1; PID:g511298
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
A;Reference number: A58370; MUID:94240111; PMID:8183893
A;Accession: S65595
A;Molecule type: mRNA
A;Residues: 487-1512,'L',1514-1522,'F',1524-1683,'V',1685-1774 <OH2>
A;Cross-references: EMBL:L22545
R;Abe, N.; Muragaki, Y.; Yoshioka, H.; Inoue, H.; Ninomiya, Y.
Biochem. Biophys. Res. Commun. 196, 576-582, 1993
A;Title: Identification of a novel collagen chain represented by extensive interruptions
A;Reference number: PN0675; MUID:94059075; PMID:8240330
A;Accession: PN0675
A;Molecule type: mRNA
A;Residues: 635-1774 <ABE>
R;Rehn, M.; Hintikka, E.; Pihlajaniemi, T.
J. Biol. Chem. 269, 13929-13935, 1994
A;Title: Primary structure of the alpha1 chain of mouse type XVIII collagen, partial str
collagen chain.
A;Reference number: A54072; MUID:94245707; PMID:8188673
A;Accession: A54072
A;Molecule type: DNA; mRNA
A;Residues: 1293-1403,'R',1405-1774 <REH3>
A;Cross-references: GB:U03714; NID:g487733; PIDN:AAA20657.1; PID:g487734
R;O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukai, N.; Vasios, G.; Lane, W.S.; Flynn, E.; Bi
Cell 88, 277-285, 1997
A;Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.
A;Reference number: A58816; MUID:97160848; PMID:9008168
A;Accession: A58816
A;Molecule type: protein
A;Residues: 1591-1610 <ORE>
A;Experimental source: hemangioendothelium cells
A;Note: inhibits endothelial cell proliferation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
lated and subsequently O-glycosylated.
C;Comment: The different splice forms of collagen alpha 1 (XVIII) may be involved in peri

---

C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of u
ay be useful in treating solid tumors.
C;Genetics:
A;Gene: MGI:Col18a1
A;Cross-references: MGI:71175
A;Map position: 10:41.0
A;Introns: 129/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 1595
A;Note: the list of introns is incomplete
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteogly
F;1-1774/Product: collagen alpha 1(XVIII) chain precursor, long splice form #status pre
F;1-239,487-1774/Product: collagen alpha 1(XVIII) chain precursor, medium splice form #
F;1-24/Domain: signal sequence #status predicted <SIG>
F;786-812/Domain: collagenous #status predicted <CO1>
F;822-896/Domain: collagenous #status predicted <CO2>
F;921-1042/Domain: collagenous #status predicted <CO3>
F;1066-1148/Domain: collagenous #status predicted <CO4>
F;1163-1204/Domain: collagenous #status predicted <CO5>
F;1218-1290/Domain: collagenous #status predicted <CO6>
F;1301-1333/Domain: collagenous #status predicted <CO7>
F;1346-1369/Domain: collagenous #status predicted <CO8>
F;1351-1353/Region: cell attachment (R-G-D) motif
F;1377-1428/Domain: collagenous #status predicted <CO9>
F;1442-1459/Domain: collagenous #status predicted <CO10>
F;1591-1774/Product: endostatin #status predicted <EST>
F;1598-1774/Region: multiplexin collagen carboxyl-terminal homologous
F;354,361,947/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match   86.0%; Score 831; DB 2; Length 1774;
Best Local Similarity   83.7%; Pred. No. 3.8e-73;
Matches 154; Conservative 17; Mismatches 13; Indels 0; Gaps 0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
               |||||||||||||||||||||||||||||||||||:||||| |||||||||||||||||
Db   1591 HTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSI 1650

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSBGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
               |||||| ||:||||||||| |||:||||| |||:::|||||||||||||:|||:|||||
Db   1651 VRRADRGSVPIVNLKDEVLSPSWDSLPSGSQGQLQPGARIFSFDGRDVLRHPAWPQKSVW 1710

Qy   121 HGSDPSGRRLTDSYCETWRTEAPAATQOASSLLAGRLLEQRAASCRHAFVVLCIENSVMT 180
               |||||||||:||||||||||||||||:||||||||||||||||||::::|||||||:||:
Db   1711 HGSDPSGRRLMESYCETWRTETTCATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMT 1770

Qy   181 SFSK 184
               ||||
Db   1771 SFSK 1774

RESULT 3
A56101
collagen alpha 1(XVIII) chain precursor, short splice form - mouse
N;Contains: endostatin
C;Species: Mus musculus (house mouse)
C;Date: 03-Oct-1995 #sequence revision 08-May-1998 #text_change 15-Sep-2003
C;Accession: A56101; A58371; S72450; S65595
R;Rehn, M.; Pihlajaniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A;Title: Identification of three N-terminal ends of type XVIII collagen chains and tiss
tif homologous to rat and Drosophila frizzled proteins.
A;Reference number: A56101; MUID:95181468; PMID:7876242
A;Accession: A56101
A;Molecule type: mRNA
A;Residues: 1-103 <REH1>
A;Cross-references: GB:U11636; NID:g618427; PIDN:AAC52178.1; PID:g618428
R;Rehn, M.; Pihlajaniemi, T.
Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994
A;Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collagenous
A;Reference number: A58371; MUID:94240112; PMID:8183894
A;Accession: A58371
A;Molecule type: mRNA
A;Residues: 1-928 <REH2>

A;Cross-references: GB:L16898; NID:g404754; PIDN:AAA37434.1; PID:g553894
R;Oh, S.P.; Kamagata, Y.; Muragata, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A;Reference number: S72450
A;Accession: S72450
A;Molecule type: mRNA
A;Residues: 28-687,'L',689-734,'F',736-751,'R',753-1315 <OHW>
R;Oh, S.P.; Kamagata, Y.; Muragata, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
A;Reference number: A58370; MUID:94240111; PMID:8183893
A;Accession: S65595
A;Molecule type: mRNA
A;Residues: 28-1315 <OHS>
A;Cross-references: EMBL:L22545
C;Genetics:
A;Gene: MGI:Col18a1
A;Cross-references: MGI:71175
A;Map position: 10:41.0
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc
F;1-25/Domain: signal sequence #status predicted <SIG>
F;24-235/Region: thrombospondin amino-terminal homologous
F;26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <M
F;327-353/Domain: collagenous #status predicted <CO1>
F;364-437/Domain: collagenous #status predicted <CO2>
F;462-583/Domain: collagenous #status predicted <CO3>
F;607-689/Domain: collagenous #status predicted <CO4>
F;704-745/Domain: collagenous #status predicted <CO5>
F;759-831/Domain: collagenous #status predicted <CO6>
F;842-874/Domain: collagenous #status predicted <CO7>
F;887-910/Domain: collagenous #status predicted <CO8>
F;892-894/Region: cell attachment (R-G-D) motif
F;918-969/Domain: collagenous #status predicted <CO9>
F;983-1000/Domain: collagenous #status predicted <CO10>
F;1132-1315/Product: endostatin #status predicted <EST>
F;1119-1315/Region: multiplexin collagen carboxyl-terminal homologous
F;1126,488/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;172-228/Disulfide bonds: #status predicted
F;240,245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;451,454,594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match   85.7%; Score 828; DB 2; Length 1315;
Best Local Similarity 83.2%; Pred. No. 5.2e-73;
Matches 153; Conservative 18; Mismatches 13; Indels 0; Gaps 0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
         ||||||| ||||||||| | ||||||||||||||||||||| ||:||||||||||||||||
Db 1132 HTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSI 1191

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
         ||||||:|||:||| :|||||: ||| |||::|||:||||||||||||||:|||| ||||
Db 1192 VRRADRGSVPIVNLKDEVLSPSWDSLPSGSGVQPGARIFSFDGRDVLRHPAWPAWPQKSVW 1251

Qy 121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
         |||||||||| :||||||| ||||||||||:|||||||||||| :::||||||||
Db 1252 HGSDPSGRRLMESYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMT 1311

Qy 181 SFSK 184
         ||||
Db 1312 SFSK 1315

RESULT 4
A53317
collagen alpha 1(XV) chain precursor - human
N;Alternate names: procollagen alpha 1(XV) chain
C;Species: Homo sapiens (man)

C;Date: 07-Jul-1995 #sequence revision 07-Jul-1995 #text_change 15-Sep-2003
C;Accession: A53317; A53146; S28778
R;Kivirikko, S.; Heinaemaeki, P.; Rehn, M.; Honkanen, N.; Myers, J.C.; Pihlajaniemi, T
J. Biol. Chem. 269, 4773-4779, 1994
A;Title: Primary structure of the alpha1 chain of human type XV collagen and exon-intro
A;Reference number: A53317; MUID:94148920; PMID:8106446
A;Accession: A53317
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1388 <KIV>
A;Cross-references: GB:L25280
A;Note: nucleotide sequence and conceptual translation not complete
R;Muragaki, Y.; Abe, N.; Ninomiya, Y.; Olsen, B.R.; Ooshima, A.
J. Biol. Chem. 269, 4042-4046, 1994
A;Title: The human alpha1(XV) collagen chain contains a large amino-terminal non-triple
A;Reference number: A53146; MUID:94140817; PMID:8307960
A;Accession: A53146
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-9, S',11-48,'V',50-94,'A',96-149,'A',151-203,'V',205-408,'A',410-569 <MUR
A;Residues: 544-640,'P',642-811,'P',813-1252 <MYE>
A;Cross-references: GB:D21230; NID:g415605; PIDN:BAA04762.1; PID:d1005294; PID:g460703
R;Myers, J.C.; Kivirikko, S.; Gordon, M.K.; Dion, A.S.; Pihlajaniemi, T.
Proc. Natl. Acad. Sci. U.S.A. 89, 10144-10148, 1992
A;Title: Identification of a previously unknown human collagen chain, alpha1(XV), chara
A;Reference number: S28778; MUID:93066196; PMID:1279671
A;Accession: S28778
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 544-640,'P',642-811,'P',813-1252 <MYE>
C;Genetics:
A;Gene: GDB:COL15A1
A;Cross-references: GDB:132578; OMIM:120325
A;Map position: 9q21-9q22
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-1388/Product: collagen alpha 1(XV) #status predicted <MAT>
F;1216-1388/Region: multiplexin collagen carboxyl-terminal homologous

Query Match   54.5%; Score 526; DB 2; Length 1388;
Best Local Similarity 56.6%; Pred. No. 2.8e-43;
Matches 99; Conservative 27; Mismatches 45; Indels 4; Gaps 1;

Qy   10 LHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGV 69
         |||:||| | | || |||||||||:||||||| :|:||||||| |||| :|||: | : |
Db 1218 LHLAALNMPFSGDIR---ADFQCFKQARAAGLLSTYRAFLSSHLQDLSTIVRKAERYSL 1273

Qy   70 PVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRR 129
         |:|||: :||| || :||| |:| :: | : |:|||::||| |:|:|||||:|| | :
Db 1274 PIVNLKGVLFNNWDSIFSGHGQQFNMHIPIYSFPDGRDIMTDPSWPQKVIWHGSSPHGVR 1333

Qy 130 LTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 184
         | | ||| |||| |:: :||| |: :: |:: || | ||:|||||:|| :||
Db 1334 LVDNYCEAWRTADTAVTGLASPLSTGKILDQKAVSCANRLIVLCIENSFMTDARK 1388

RESULT 5
T22002
hypothetical protein F39H11.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T22002
R;White, S.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19500
A;Accession: T22002
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-650 <WIL>
A;Cross-references: EMBL:Z81079; PIDN:CAB03084.1; GSPDB:GN00019; CESP:F39H11.4
A;Experimental source: clone F39H11
C;Genetics:
A;Gene: CESP:F39H11.4
A;Map position: 1
A;Introns: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

```
Query Match              37.9%;   Score 366; DB 2; Length 650;
Best Local Similarity    44.1%;   Pred. No. 6.2e-28;
Matches   75;  Conservative   26;  Mismatches   63;  Indels   26;  Gaps    4;

Qy    9 VLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTG 68
          |:|:|:||:|  :::|||||||||||||||||||||||||||||||:||||||||||||
Db  464 VIHMIALSQPFSGNLHGLRGADLQCYREARAAGYTTFRAMLSSNVQDLVRIVHSVD-FD 522

Qy   69 VPVVNLRDEVLFPSWEALFSGSEGQLKPGCARIPSFDGRDVLQHPAWPRKSVWHGSDPSGR 128
          ||||||||||||||||||||||||||   | |:|||||||||||||||||||||||||||
Db  523 TTVVNVAGHHLFPSWRSFVNGA--QMNPHAKLFSFDRHDVLNDSRWPDKRVWHGSKDGGI 580

Qy  129 RLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAAS--CRHAFVVLCIEN 176
          |||||||||||||||||||||||:||||||||:||    :|||||:|
Db  581 R-AEQYCDGWRRADSSLITSLAGHISSNTSIFQSSGSEKCENKLVVLCVEN 629
```

RESULT 6

A83204

alginate biosynthesis protein Alg8 PA3541 [imported] - Pseudomonas aeruginosa (strain PA...

C;Species: Pseudomonas aeruginosa
C;Date:15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83204
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br...
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,...
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathc...
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A83204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-494 <STO>
A;Cross-references: GB:AE004774; GB:AE004091; NID:g9945684; PIDN:AAG06929.1; GSPDB:GN001...
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA3541

```
Query Match               9.0%;   Score 87; DB 2; Length 494;
Best Local Similarity    24.6%;   Pred. No. 1.2;
Matches   45;  Conservative   20;  Mismatches   68;  Indels   50;  Gaps    7;

Qy    4 QDFQLVLHLVALNSPQPGGMRGIRGADF----------QCFQQARAAGLAGTFRAFLSSR 53
          :||: ||| ||: |  ||| ||   :      |     : |||||||||||||||||||||
Db   39 KDFILLIGAVGIWRYSMGGVHFLRGMLFLHVVYPYYRRRVRQLGSAADPSHVFLMVTSFR 98

Qy   54 LQDL-------YSIVRRADRTGVP------VVNLRDEVLFPS-WEAL------------ 86
          |:||        :|:||||  ||          ::||||||| : ||
Db   99 IDALITAMVYRSVIREAIDSGYPTTVVCSIVEMSDEYLVRSLWEKNPPDRVSLDFVRIP 158

Qy   87 FSGSEGQLKPGARIFS-----------FDGRDVLQHPAWPRKSVWHGSDPS-GRRLTDS 133
          ::|: |    ||              |||||||||||||||||||||  |||||
Db  159 GTGKRDGLAYGFRAISRHLPDDDAVVAVIDGDTVLDHGVVKKTVPWFKLFPNVGGLTTNE 218

Qy  134 YCE 136
          :||
Db  219 FCE 221
```

RESULT 7

H59432

RhoGAP protein homolog [imported] - human

C;Species: Homo sapiens (man)
C;Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 23-Sep-2002
C;Accession: H59432
R;Rhodes, S.
submitted to GenBank, October 2001
A;Description: Novel human gene mapping to chromosome 13, similar to rat RhoGAP.
A;Reference number: H59432
A;Accession: H59432
A;Status: preliminary
A;Molecule type: DNA

A;Residues: 1-995 <RHO>
A;Cross-references: GB:NP_443083; PID:g16445031; PIDN:NP_443083.1

```
Query Match               8.5%;   Score 82; DB 2; Length 995;
Best Local Similarity    23.0%;   Pred. No. 8.9;
Matches   37;  Conservative   26;  Mismatches   52;  Indels   46;  Gaps    7;

Qy   52 SRLQDLYSIVRRADRTGVP----------------VVNLRDEVLFPSWEALFSGSEGQLKPG 97
          ||:||||||||||||||                 |::|||||||||||:||:::|||:||
Db   36 SRVDLYTLLPRGDRNGSPGGTGMRNTTSESVLTDLSEPEVCSIHSBSSGGSDSRSQPG 95

Qy   98 ---------ARIFSFD----GRDVLQHPAWPRKSVWHGSDPSGRRLTDSY---CETWR 139
                    |:|:|      ||||||||||||||||||||||| :   :|||
Db   96 QCCTDNPVMLDAPLVSSLPQPPRDVLNHPFHPK------NEKPTRARAKSFLKRMETLR 149

Qy  140 TEAPAATGQASS-----LLAGRLLEQEAASCRHAFVVLCIE 175
          :    :   :||      ||  :|   :  |:|||||: :|
Db  150 GKGAHGRHKGSGRTGGLVISGPMLQQEPESFK---AMQCIQ 187
```

RESULT 8

T36950

probable transcription regulator - Streptomyces coelicolor

C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T36950
R;Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.l
submitted to the EMBL Data Library, September 1999
A;Reference number: Z21607
A;Accession: T36950
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-190 <SEE>
A;Cross-references: EMBL:AL109962; PIDN:CAB53135.1; GSPDB:GN00070; SCOEDB:SCJ1.17
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCJ1.17

```
Query Match               8.4%;   Score 81.5;  DB 2;  Length 190;
Best Local Similarity    25.1%;   Pred. No. 1.4;
Matches   46;  Conservative   18;  Mismatches   60;  Indels   59;  Gaps   10;

Qy   11 HLVALNSPQPGGMRGIRGADF----------------------QCFQQARAAGLAGTFRAFLSSR 53
          ||:||::| |::| |||||                        |||||||||||||||||||||
Db   18 HLAAARATAPDGPREVEGKETMTSTTLPIAMPAQHRERLLRLAREVSFPPGARLFEEDR 77

Qy   54 LQDLYSIVRRADRTGVPVVNLR----------------DEVLFPSW3----ALFSGSEGQLK 95
          |:||| :|||||| ||:|                    ||||:|||        ||||||||
Db   78 PADRFWIV----RTGTVVLDIRVPDGRNAAVESLGHGSLLGWSWHPPYRWHQGAEA-LS 132

Qy   96 PGARIFSFDGRDVL----OHPAWPRK-SVWHGSDPSGR------RLTDSYCETWRTEAPA 144
          |||||:||||||     |||||||| ||||||||||        ||||||||:
Db  133 P-VRAWEFDAKTVLTTCAEHPEFGRAIASWVGQVVADRLQKTRIRLLDLY------APH 184

Qy  145 ATG 147
          :||
Db  185 GSG 187
```

RESULT 9

G90704

probable Vgr protein [imported] - Escherichia coli (strain O157:H7, substrain RIMD 050'...

C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: G90704
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C...
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and ge...
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: G90704
A;Status: preliminary
A;Molecule type: DNA

```
A;Residues: 1-633 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB34030.1; PID:g13360065; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0607

Query Match           8.4%;  Score 81.5;  DB 2;  Length 633;
Best Local Similarity   24.4%;  Pred. No. 5.8;
Matches  40; Conservative  18; Mismatches  59; Indels  47; Gaps   8;

Qy      3 HQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYISIVR  62
            |||:|                |  :  :/:|
Db    247 HQDYQRTQYEV---YDYPGRFKGAHGQNFARWQMEGWRNNAETARGM--SRSPEIWP-GR 300

Qy     63 RADRTGVPVVNLRDEVLFPSWEALFS---GSEGQLKPGARIFSFDGRDVLQHPAWPRKSV 119
            :  :  :                            ||||
Db    301 RIVLTGHPQANLNRE-----WQVVASELHGEQPQAVPGRR------------------ 335

Qy    120 WHGSDPSGRRLTDSYC-----ETWRTE--APAATGQASSLLAG 155
            |||:  :||                |||      |:  :
Db    336 -----GAGTALENHFAVIPADRTWRPQPRLKPLVDGPQSAVVTG 374


RESULT 10
B85555
probable Vgr protein [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Nov-2001
C;Accession: B85555
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iler, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 406, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: B85555
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-633 <STO>
A;Cross-references: GB:AE005174; NID:g12513451; PIDN:AAG54902.1; GSPDB:GN00145; UWGP:Z07
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Accession: Z07707

Query Match           8.4%;  Score 81.5;  DB 2;  Length 633;
Best Local Similarity   24.4%;  Pred. No. 5.8;
Matches  40; Conservative  18; Mismatches  59; Indels  47; Gaps   8;

Qy      3 HQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYISIVR  62
            |||:|                |  :  :/:|
Db    247 HQDYQRTQYEV---YDYPGRFKGAHGQNFARWQMEGWRNNAETARGM--SRSPEIWP-GR 300

Qy     63 RADRTGVPVVNLRDEVLFPSWEALFS---GSEGQLKPGARIFSFDGRDVLQHPAWPRKSV 119
            :  :  :                            ||||
Db    301 RIVLTGHPQANLNRE-----WQVVASELHGEQPQAVPGRR------------------ 335

Qy    120 WHGSDPSGRRLTDSYC-----ETWRTE--APAATGQASSLLAG 155
            |||:  :||                |||      |:  :
Db    336 -----GAGTALENHFAVIPADRTWRPQPRLKPLVDGPQSAVVTG 374


RESULT 11
A55461
NAD(P)-arginine ADP-ribosyltransferase (EC 2.4.2.31) AT1 - chicken
C;Species: Gallus gallus (chicken)
C;Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 03-Jun-2002
C;Accession: A55461
R;Tsuchiya, M.; Hara, N.; Yamada, K.; Osago, H.; Shimoyama, M.
J. Biol. Chem. 269, 27451-27457, 1994
A;Title: Cloning and expression of cDNA for arginine-specific ADP-ribosyltransferase frc
A;Reference number: A55461; MUID:95050487; PMID:7961658
A;Accession: A55461
A;Status: preliminary
A;Molecule type: mRNA
```

```
A;Residues: 1-312 <TSU>
A;Cross-references: GB:D31864; NID:g1638784; PIDN:BAA06664.1; PID:g1638785
C;Superfamily: NAD(P)+arginine ADP-ribosyltransferase
C;Keywords: glycosyltransferase; NAD; pentosyltransferase

Query Match           8.4%;  Score 81;  DB 2;  Length 312;
Best Local Similarity   26.3%;  Pred. No. 2.8;
Matches  50; Conservative  13; Mismatches  51; Indels  76; Gaps  12;

Qy      2 THQD--------FQLVLHLVALNSPQP-----GGMRGIRGADFQCFQQARAAGLLAG  44
            :|||               |||||||
Db    128 SHQDYIHSYHFKTLHFFLTQALFALRASQPRCYVYRGVRGIR------FMTQR--GKSV 179

Qy     45 TFRAFLSSRL---------QDLYSIVRRADRTGVPVVNL-----RDEVLFPSWEAL---- 86
            |||||||           :|  |||  ||||||||
Db    180 RFGQFTSTSLRKEATVNFGQDTLFVVKTC--YGVPIKQFSFFPSEDEVLIPPEVFEVIN 237

Qy     87 FSGSBGQ-------------LKPGARIFSFDGRDV------LQHPAWP-RKSVW 120
            :                :                    :  :
Db    238 FSNDRGSVKIQLHSKGKMSTHNCELLKFQGGQWGRGHQEVGLGSPGLSLPVLPCRRRVW 297

Qy    121 HG-----SDP 125
            ||
Db    298 EGLGHREGDP 307


RESULT 12
G83127
conserved hypothetical protein PA4149 [imported]  - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: G83127
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic path
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: G83127
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-346 <STO>
A;Cross-references: GB:AE004831; GB:AE004091; NID:g9950347; PIDN:AAG07536.1; GSPDB:GN00
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA4149

Query Match           8.2%;  Score 79.5;  DB 2;  Length 346;
Best Local Similarity   27.6%;  Pred. No. 4.4;
Matches  45; Conservative  19; Mismatches  56; Indels  43; Gaps   9;

Qy     17 SPQPG-GMRGIRGAD--------FQCFQQARAAGLAGTFRAFLSSRLQDLYISIVRRADRTGVP  70
            |||:       ||||||
Db    198 SPQPFIGARAISRAEDLAEVFVSPAEPHAIGL---------SALCGLWSPVSREEPHGAW 248

Qy     71 V---VNLRDEVLFPSWEALPSG-----SEGQLKPG-------ARIFSFDGRDVLQHPAWP 115
            :                |||      :|:|
Db    249 MRLHPDARESLLVPLAPGLLVGCGVSAAGYLQPGVAHAPSLSSGTLALDGERETEFSATD 308

Qy    116 RKSVWHGSDPSGRRLTDSYCETWRTEAPAATGQAS--SLLAGR 156
            |||:                              |||||
Db    309 RPSI--TLDPSG---------PFSVDVPATLAYAARHRLLAGQ 340


RESULT 13
D90658
VgrG protein [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: D90658
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
```

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gend
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D90658
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-713 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33659.1; PID:g13359692; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0236

```
Query Match       8.2%;  Score 79.5;  DB 2;  Length 713;
Best Local Similarity  24.4%;  Pred. No. 10;
Matches   40;  Conservative  18;  Mismatches   59;  Indels  47;  Gaps   8;

Qy    3 HQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVR 62
          || |||  : :   ||  :|  :|:   ||  :    |||:|||  |:|  ||:
Db  247 HQDYQRTQYEV---YDYPGRFKGAHGQNFARWQMDGWRNNAEVARG--TSRSPEIWP-GR 300

Qy   63 RADRTGVPVVNLRDEVLFPSWEALFS---GSEGQLKPGARIFSFDGRDVLQHPAWPRKSV 119
          | : |||  ||  :|
Db  301 RIVLTGHPQANLNRE----WQVASELHGEQPQAVPGRR---------------- 335

Qy  120 WHGSDPSGRRLTDSYC-----ETWRTE--APAATGQASSLLAG 155
          ||:    |  |    :::  |  |:|:
Db  336 -----GSGTTLDNHFAVIPADRTWRPQPLLKPLVDGPQSAVVTG 374
```

RESULT 14
D85509
hypothetical protein Z0267 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: D85509
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D85509
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-713 <STO>
A;Cross-references: GB:AE005174; NID:g12512976; PIDN:AAG54536.1; GSPDB:GN00145; UWGP:202
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0267

```
Query Match       8.2%;  Score 79.5;  DB 2;  Length 713;
Best Local Similarity  24.4%;  Pred. No. 10;
Matches   40;  Conservative  18;  Mismatches   59;  Indels  47;  Gaps   8;

Qy    3 HQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVR 62
          || |||  : :   ||  :|  :|:   ||  :    |||:|||  |:|  ||:
Db  247 HQDYQRTQYEV---YDYPGRFKGAHGQNFARWQMDGWRNNAEVARG--TSRSPEIWP-GR 300

Qy   63 RADRTGVPVVNLRDEVLFPSWEALFS---GSEGQLKPGARIFSFDGRDVLQHPAWPRKSV 119
          | : |||  ||  :|
Db  301 RIVLTGHPQANLNRE----WQVASELHGEQPQAVPGRR---------------- 335

Qy  120 WHGSDPSGRRLTDSYC-----ETWRTE--APAATGQASSLLAG 155
          ||:    |  |    :::  |  |:|:
Db  336 -----GSGTTLDNHFAVIPADRTWRPQPLLKPLVDGPQSAVVTG 374
```

RESULT 15
T36083
hypothetical protein SCE134.01c - Streptomyces coelicolor (fragment)
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T36083
R;Saunders, D.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, April 1999

A;Reference number: Z21596
A;Accession: T36083
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-672 <SAU>
A;Cross-references: EMBL:AL049661; PIDN:CAB41199.1; GSPDB:GN00070; SCOEDB:SCE134.01c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCE134.01c

```
Query Match       8.2%;  Score 79;  DB 2;  Length 672;
Best Local Similarity  31.5%;  Pred. No. 11;
Matches   39;  Conservative  10;  Mismatches   35;  Indels  40;  Gaps   8;

Qy   48 AFLSSRLQDLYSIVRRADRTGVPVVNLRDEVLFPSWEALFSSGEGQLKPGARIFSFDGRD 107
          |  ||| |  |    |||  ||                               ::
Db  465 AMFSERLG-----VQRVDRTLV-VTPLGRVVILDELEA----------EGPREWTY---- 504

Qy  108 VLQHPAWPRKSVWHGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEDEAASCRH 167
          :||  :|||       |:|      |:      |  |:|                  ||
Db  505 -LLHSDWPTTQL----DPDG--------WRLE----SGPASARLT-RLLPTDTATATRH 544

Qy  168 AFVV 171
          ||
Db  545 RTVV 548
```

Search completed: March 26, 2004, 13:38:40
Job time : 14.4444 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        March 26, 2004, 13:29:16 ; Search time 9.33333 Seconds
                                            (without alignments)
                                          1026.526 Million cell updates/sec

Title:         US-09-938-391-4
Perfect score: 966
Sequence:      1 HTHQDFQLVLHLVALNSPQP.........CRHAFVVLCIENSVMTSFSK 184

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:          141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     SwissProt_42:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.

                              SUMMARIES

                    %
Result        Query
No.   Score Match Length DB  ID              Description
----- ----- ----- ------ --  --              -----------
1     835   86.4  1516   1   CAIH_HUMAN      P39060 homo sapien
2     831   86.0  1774   1   CAIH_MOUSE      P39061 mus musculu
3     526   54.5  1388   1   CAIE_HUMAN      P39059 homo sapien
4     87    9.0   494    1   ALG8_PSEAE      Q52463 pseudomonas
5     85.5  8.9   556    1   GLI_CHICK       P55878 gallus gall
6     82    8.5   995    1   SRI3_HUMAN      Q9y3m8 homo sapien
7     82    8.5   1233   1   NME3_HUMAN      Q14957 homo sapien
8     81    8.4   312    1   NRT1_CHICK      P55806 gallus gall
9     77    8.0   419    1   P47K_PSECL      P31521 pseudomonas
10    76.5  7.9   953    1   CAR4_MOUSE      Q8bhb0 mus musculu
11    76    7.9   7073   1   R1AB_CVHSA      P59641 h replicase
12    75.5  7.8   512    1   PPX_ECOLI       P29014 escherichia
13    73.5  7.6   326    1   PIM3_HUMAN      Q86v86 homo sapien
14    73.5  7.6   807    1   YCDS_ECOLI      P75907 escherichia
15    72    7.5   646    1   NA95_HUMAN      Q9ulx6 homo sapien
16    72    7.5   309    1   YHCC_ECOLI      P45476 escherichia
17    72    7.5   5217   1   HTS1_COCCA      Q01886 cochliobolu
18    71.5  7.4   356    1   MURG_YERPE      Q8zie9 yersinia pe
19    71.5  7.4   1848   1   CCAA_DROME      P91645 drosophila
20    71    7.3   999    1   MERK_HUMAN      Q12866 homo sapien
21    70.5  7.3   312    1   NRT2_CHICK      P55807 gallus gall
22    70.5  7.3   317    1   GGH_RAT         Q62867 rattus norv
23    70    7.2   185    1   NPW_RAT         Q8k1m5 rattus norv
24    69.5  7.2   300    1   NARE_CHICK      Q92080 gallus gall
25    69.5  7.2   415    1   VE2_PAPVE       P11329 european el
26    69.5  7.2   2799   1   EDD_HUMAN       O95071 homo sapien
27    69    7.2   676    1   UL06_HSV11      P10190 herpes simp
28    69    7.1   838    1   GLGB_STRAW      Q82jf0 streptomyce
29    68.5  7.1   319    1   HA11_RAT        P15978 rattus norv
30    68.5  7.1   326    1   PIM3_MOUSE      P58750 mus musculu
31    68.5  7.1   326    1   PIM1_RAT        O70444 rattus norv
32    68.5  7.1   428    1   NER3_BOVIN      O97859 bos taurus
33    68.5  7.1   739    1   VNUC_EBOZ5      O72142 ebola virus
34          68.5  7.1   739    1   VNUC_EBOZM      P18272 ebola virus
35          68.5  7.1   843    1   DPOL_HPBVJ      P17394 hepatitis b
36          68    7.0   303    1   Y247_HUMAN      Q92537 homo sapien
37          68    7.0   663    1   DYIN_DROME      Q24246 drosophila
38          68    7.0   1722   1   LY75_MESAU      Q920p9 mesocricetu
39          67.5  7.0   273    1   FDHD_PASMU      P57842 pasteurella
40          67.5  7.0   436    1   PNCB_VIBPA      Q87je4 vibrio para
41          67.5  7.0   448    1   HGD_BRAJA       Q89xh1 bradyrhizob
42          67.5  7.0   477    1   AMT_MYCTU       Q10968 mycobacteri
43          67.5  7.0   640    1   PPCM_HUMAN      Q16822 homo sapien
44          67.5  7.0   657    1   MY16_MOUSE      P17564 mus musculu
45          67.5  7.0   1327   1   YNCA_CAEEL      P34544 caenorhabdi

                              ALIGNMENTS

RESULT 1
CAIH_HUMAN
ID   CAIH_HUMAN     STANDARD;      PRT;   1516 AA.
AC   P39060; Q9UK38, Q9Y6Q7; Q9Y6Q8;
DT   01-FEB-1995 (Rel. 31, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin].
GN   COL18A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98164096; PubMed=9503365;
RA   Saarela J., Ylikarppa R., Rehn M.,  Purmonen S., Pihlajaniemi T.;
RT   "Complete primary structure of two variant forms of human type XVIII
RT   collagen and tissue-specific differences in the expression of the
RT   corresponding transcripts.";
RL   Matrix Biol. 16:319-328(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20289799; PubMed=10830953;
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA   Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
RA   Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA   Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mituyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
RN   [3]
RP   SEQUENCE OF 834-1516 FROM N.A.
RX   MEDLINE=94245237; PubMed=8188291;
RA   Oh S.P., Warman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,
RA   Olsen B.R.;
RT   "Cloning of cDNA and genomic DNA encoding human type XVIII collagen
RT   and localization of the alpha 1(XVIII) collagen gene to mouse
RT   chromosome 10 and human chromosome 21.";
RL   Genomics 19:494-499(1994).
RN   [4]
RP   SEQUENCE OF 1334-1516 FROM N.A.
RC   TISSUE=Placenta;
RA   Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.;
RT   "Cloning and expression of human endostatin gene in Escherichia
RT   coli.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   INVOLVEMENT IN KNOBLOCH SYNDROME.
```

RX   MEDLINE=20400145; PubMed=10942434;
RA   Sertie A.L., Sossi V., Camargo A.A., Zatz M., Brahe C.,
RA   Passos-Bueno M.R.;
RT   "Collagen XVIII, containing an endogenous inhibitor of angiogenesis
RT   and tumor growth, plays a critical role in the maintenance of retinal
RT   structure and in neural tube closure.";
RL   Hum. Mol. Genet. 9:2051-2058(2000).
RN   [6]
RP   VARIANT ASN-1437.
RX   MEDLINE=21518361; PubMed=11606364;
RA   Iughetti P., Suzuki O., Godoi P.H., Alves V.A., Sertie A.L.,
RA   Zorick T., Soares F., Camargo A., Moreira E.S., di Loreto C.,
RA   Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;
RT   "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes
RT   for the development of prostatic adenocarcinoma.";
RL   Cancer Res. 61:7375-7378(2001).
CC   -!- FUNCTION: COL18A1 probably plays a major role in determining the
CC   retinal structure as well as in the closure of the neural tube.
CC   -!- FUNCTION: Endostatin potently inhibits endothelial cell
CC   proliferation and angiogenesis. May inhibit angiogenesis by
CC   binding to the heparan sulphate proteoglycans involved in growth
CC   factor signalling.
CC   -!- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=2;
CC   Name=Long; Synonyms=NC-493;
CC   IsoId=P39060-1; Sequence=Displayed;
CC   Name=Short; Synonyms=NC1-303;
CC   IsoId=P39060-2; Sequence=VSP_001155, VSP_001156;
CC   -!- TISSUE SPECIFICITY: Present in multiple organs with highest levels
CC   in liver, lung and kidney.
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC   unit (G-x-Y) are hydroxylated in some or all of the chains.
CC   -!- POLYMORPHISM: There is an association between a polymorphism in
CC   position 1437 and prostate cancer. Heterozygous Asn-1437
CC   individuals have a 2.5 times increased chance of developing
CC   prostate cancer as compared with homozygous Asp-1437 individuals.
CC   -!- DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KNO)
CC   [MIM:267750]; an autosomal recessive disorder defined by the
CC   occurrence of high myopia, vitreoretinal degeneration with retinal
CC   detachment, macular abnormalities and occipital encephalocele.
CC   -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC   INTERRUPTED HELICES (FACIT) FAMILY.
CC   -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AF018081; AAC39658.1; -.
DR   EMBL; AF018082; AAC39659.1; -.
DR   EMBL; AL163302; CAB90482.1; -.
DR   EMBL; L22548; AAA51864.1; -.
DR   EMBL; AF184060; AAF01310.1; ALT_INIT.
DR   PDB; 1BNL; 02-DEC-98.
DR   GlycoSuiteDB; P39060; -.
DR   Genew; HGNC:2195; COL18A1.
DR   MIM; 120328; -.
DR   MIM; 267750; -.
DR   GO; GO:0005581; C:collagen; TAS.
DR   GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR   GO; GO:0008285; P:negative regulation of cell proliferation; TAS.
DR   GO; GO:0007601; P:vision; TAS.
DR   InterPro; IPRO08161; Clg_helix.
DR   InterPro; IPRO08160; Collagen.
DR   InterPro; IPRO08995; ConA_like_lec_gl.
DR   InterPro; IPRO01791; Laminin_G.
DR   InterPro; IPRO03129; TSPN.
DR   Pfam; PF01391; Collagen; 7.
DR   Pfam; PF02210; TSPN; 1.

DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
KW   Polymorphism; 3D-structure.
FT   SIGNAL   1   23   POTENTIAL.
FT   CHAIN   24   1516   COLLAGEN ALPHA 1(XVIII) CHAIN.
FT   CHAIN   1334   1516   ENDOSTATIN.
FT   DOMAIN   221   409   TSP N-TERMINAL.
FT   DOMAIN   410   516   NONHELICAL REGION 1 (NC1).
FT   DOMAIN   517   550   TRIPLE-HELICAL REGION 1 (COL1).
FT   DOMAIN   551   560   NONHELICAL REGION 2 (NC2).
FT   DOMAIN   561   640   TRIPLE-HELICAL REGION 2 (COL2).
FT   DOMAIN   641   664   NONHELICAL REGION 3 (NC3).
FT   DOMAIN   665   786   TRIPLE-HELICAL REGION 3 (COL3).
FT   DOMAIN   787   809   NONHELICAL REGION 4 (NC4).
FT   DOMAIN   810   892   TRIPLE-HELICAL REGION 4 (COL4).
FT   DOMAIN   893   906   NONHELICAL REGION 5 (NC5).
FT   DOMAIN   907   948   TRIPLE-HELICAL REGION 5 (COL5).
FT   DOMAIN   949   961   NONHELICAL REGION 6 (NC6).
FT   DOMAIN   962   1034   TRIPLE-HELICAL REGION 6 (COL6).
FT   DOMAIN   1035   1044   NONHELICAL REGION 7 (NC7).
FT   DOMAIN   1045   1077   TRIPLE-HELICAL REGION 7 (COL7).
FT   DOMAIN   1078   1089   NONHELICAL REGION 8 (NC8).
FT   DOMAIN   1090   1111   TRIPLE-HELICAL REGION 8 (COL8).
FT   DOMAIN   1112   1118   NONHELICAL REGION 9 (NC9).
FT   DOMAIN   1119   1173   TRIPLE-HELICAL REGION 9 (COL9).
FT   DOMAIN   1174   1186   NONHELICAL REGION 10 (NC10).
FT   DOMAIN   1187   1204   TRIPLE-HELICAL REGION 10 (COL10).
FT   DOMAIN   1205   1516   NONHELICAL REGION 11 (NC11).
FT   CARBOHYD   68   68   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   129   129   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   164   164   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   691   691   O-LINKED (GALNAC...) (POTENTIAL).
FT   CARBOHYD   1329   1329   O-LINKED (GALNAC...).
FT   /FTId=CAR_000150.
FT   DISULFID   1366   1506   BY SIMILARITY.
FT   DISULFID   1468   1498   BY SIMILARITY.
FT   SITE   1095   1097   CELL ATTACHMENT SITE (POTENTIAL).
FT   VARSPLIC   1   180   Missing (in isoform Short).
FT   /FTId=VSP_001155.
FT   VARSPLIC   181   215   HTTEAGTLPAPTPSPPSLGRPWAPLTGPSVPPSS -> MA
FT   PRCPWPWPRRRLLDVLAPLVLLLGVRAASAEP (in
FT   isoform Short).
FT   /FTId=VSP_001156.
FT   VARIANT   1437   1437   D -> N (increased risk of developing
FT   prostate cancer).
FT   /FTId=VAR_012709.
FT   CONFLICT   428   428   F -> S (IN REF. 2).
FT   CONFLICT   841   841   I -> V (IN REF. 2).
FT   CONFLICT   877   877   V -> L (IN REF. 3).
FT   CONFLICT   886   886   P -> R (IN REF. 3).
FT   CONFLICT   912   912   P -> R (IN REF. 3).
FT   CONFLICT   933   933   R -> L (IN REF. 3).
FT   CONFLICT   975   975   P -> L (IN REF. 3).
FT   CONFLICT   1064   1064   A -> P (IN REF. 3).
FT   CONFLICT   1084   1084   L -> K (IN REF. 3).
FT   CONFLICT   1120   1120   P -> A (IN REF. 3).
FT   CONFLICT   1123   1123   P -> A (IN REF. 3).
FT   CONFLICT   1126   1126   P -> PPGP (IN REF. 2).
FT   CONFLICT   1206   1206   G -> GQ (IN REF. 3).
FT   CONFLICT   1304   1304   R -> G (IN REF. 3).
FT   CONFLICT   1314   1314   A -> G (IN REF. 3).
FT   CONFLICT   1323   1324   LR -> CG (IN REF. 3).
FT   CONFLICT   1443   1443   R -> T (IN REF. 4).
FT   CONFLICT   1483   1483   S -> Y (IN REF. 4).
SQ   SEQUENCE   1516 AA; 153840 MW; 3C70F29A4476EB76 CRC64;

Query Match   86.4%; Score 835; DB 1; Length 1516;
Best Local Similarity   85.1%; Pred. No. 2.2e-72;
Matches 154; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

```
Qy    1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
        |:|:|| |||||||||| |||||||||||||||||||||||||||||||||||||||||
Db 1334 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 1393

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSSEQLKPGARIPSFDGRDVLQHPAWPRKSVW 120
        |||||| |:|:||||:|:|||||||||||||||||||||||:|:|| ||||||| ||:
Db 1394 VRRADRAAVPIVNLKDELLFPSWEALFSGSSESPLKPGARIPSFDGKDVLRHPTWPQKSVW 1453

Qy  121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRILEGBAASCRHAFVVLCIENSVMT 180
        |||||:|||||:||||||||||||:||||||||||:|||| ||||:||:||||||||:||
Db 1454 HGSDPNGRRLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 1513

Qy  181 S 181
        :
Db 1514 A 1514

RESULT 2
CA1H_MOUSE
ID  CA1H_MOUSE    STANDARD;    PRT;  1774 AA.
AC  P39061; Q60672; Q61437; Q62001; Q62002; Q9JK63;
DT  01-FEB-1995 (Rel. 31, Created)
DT  10-OCT-2003 (Rel. 42, Last sequence update)
DT  10-OCT-2003 (Rel. 42, Last annotation update)
DE  Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin].
GN  COL18A1.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A. (ISOFORM 3).
RC  STRAIN=BALB/c; TISSUE=Liver;
RX  MEDLINE=94245707; PubMed=8188673;
RA  Rehn M.V., Hintikka E., Pihlajaniemi T.;
RT  "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
RT  partial structure of the corresponding gene, and comparison of the
RT  alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
RT  chain.";
RL  J. Biol. Chem. 269:13929-13935(1994).
RN  [2]
RP  SEQUENCE FROM N.A. (ISOFORMS 1, 2 AND 3).
RX  MEDLINE=96435922; PubMed=8838808;
RA  Rehn M., Hintikka E., Pihlajaniemi T.;
RT  "Characterization of the mouse gene for the alpha-1 chain of type
RT  XVIII collagen (COL18A1) reveals that the three variant N-terminal
RT  polypeptide forms are transcribed from two widely separated
RT  promoters.";
RL  Genomics 32:436-446(1996).
RN  [3]
RP  SEQUENCE OF 1-1387 FROM N.A. (ISOFORM 3).
RX  MEDLINE=94240112; PubMed=8183894;
RA  Rehn M.V., Pihlajaniemi T.;
RT  "Alpha 1(XVIII), a collagen chain with frequent interruptions in the
RT  collagenous sequence, a distinct tissue distribution, and homology
RT  with type XV collagen.";
RL  Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RN  [4]
RP  SEQUENCE OF 487-1774 FROM N.A.
RC  TISSUE=Liver;
RX  MEDLINE=94240111; PubMed=8183893;
RA  Oh S.P., Kamagata Y., Muragaki Y., Timmons S., Ooshima A., Olsen B.R.;
RT  "Isolation and sequencing of cDNAs for proteins with multiple domains
RT  of Gly-Xaa-Yaa repeats identify a distinct family of collagenous
RT  proteins.";
RL  Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).
RN  [5]
RP  SEQUENCE OF 1591-1774 FROM N.A.
RX  MEDLINE=21217748; PubMed=11321448;
RA  Jia S., Zhu F., Li H., He F., Xiu R.-J.;
RT  "Anticancer treatment of endostatin gene therapy by targeting tumor
RT  neovasculature in C57/BL mice.";
RL  Clin. Hemorheol. Microcirc. 23:251-257(2000).
RN  [6]
RP  CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.
RX  MEDLINE=97160848; PubMed=9008168;
RA  O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vasios G., Lane W.S.,
RA  Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;
RT  "Endostatin: an endogenous inhibitor of angiogenesis and tumor
RT  growth.";
RL  Cell 88:277-285(1997).
RN  [7]
RP  X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.
RX  MEDLINE=98169382; PubMed=9501087;
RA  Hohenester E., Sasaki T., Olsen B.R., Timpl R.;
RT  "Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A
RT  resolution.";
RL  EMBO J. 17:1656-1664(1998).
CC  -!- FUNCTION: Endostatin potently inhibits endothelial cell
CC      proliferation and angiogenesis. May inhibit angiogenesis by
CC      binding to the heparan sulfate proteoglycans involved in growth
CC      factor signaling.
CC  -!- ALTERNATIVE PRODUCTS:
CC      Event=Alternative promoter;
CC        Comment=2 isoforms, 1 (shown here) and 3, are produced by use
CC        of alternative promoters;
CC      Event=Alternative splicing; Named isoforms=3;
CC      Name=1; Synonyms=NC1-764;
CC        IsoId=P39061-3; Sequence=Displayed;
CC      Name=2; Synonyms=Long, NC1-517;
CC        IsoId=P39061-1; Sequence=VSP_008303;
CC        Note=Produced by alternative splicing of isoform 1;
CC      Name=3; Synonyms=Short, NC1-301;
CC        IsoId=P39061-2; Sequence=VSP_001157, VSP_001158;
CC  -!- PTM: Prolines at the third position of the tripeptide repeating
CC      unit (G-X-Y) are hydroxylated in some or all of the chains.
CC  -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC      INTERRUPTED HELICES (FACIT) FAMILY.
CC  -!- SIMILARITY: Contains 1 frizzled (FZ) domain.
CC  -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC  ---------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  ---------------------------------------------------------------------
DR  EMBL; L16898; AAA37434.1; -.
DR  EMBL; U03714; AAA20657.1; -.
DR  EMBL; U03715; AAC52901.1; -.
DR  EMBL; U34606; AAC52901.1; JOINED.
DR  EMBL; U34608; AAC52901.1; JOINED.
DR  EMBL; U34609; AAC52901.1; JOINED.
DR  EMBL; U34610; AAC52901.1; JOINED.
DR  EMBL; U34611; AAC52901.1; JOINED.
DR  EMBL; U34612; AAC52901.1; JOINED.
DR  EMBL; U34613; AAC52901.1; JOINED.
DR  EMBL; U03716; AAC52901.1; JOINED.
DR  EMBL; U03718; AAC52901.1; JOINED.
DR  EMBL; U03715; AAC52902.1; -.
DR  EMBL; U34607; AAC52902.1; JOINED.
DR  EMBL; U34608; AAC52902.1; JOINED.
DR  EMBL; U34609; AAC52902.1; JOINED.
DR  EMBL; U34610; AAC52902.1; JOINED.
DR  EMBL; U34611; AAC52902.1; JOINED.
DR  EMBL; U34612; AAC52902.1; JOINED.
DR  EMBL; U34613; AAC52902.1; JOINED.
DR  EMBL; U03716; AAC52902.1; JOINED.
DR  EMBL; U03718; AAC52902.1; JOINED.
DR  EMBL; U03715; AAC52903.1; -.
DR  EMBL; U03716; AAC52903.1; JOINED.
DR  EMBL; U03718; AAC52903.1; JOINED.
DR  EMBL; U34608; AAC52903.1; JOINED.
```

```
Query Match
Best Local Similarity   86.0%;   Score 831;   DB 1;   Length 1774;
                        83.7%;  Pred. No. 6.4e-72;
Matches  154;  Conservative  17;  Mismatches  13;  Indels  0;  Gaps  0;

Qy     1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI   60
           |||||||||||| |||||||||||||||||||||||||||||||||:|||||||||||||
Db  1591 HTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSI 1650

Qy    61 VRRADRTGPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFPGRDVLQHPAWPRKSVW  120
           |||||||:|:|||||||||:|||:|||||||:||||:|||||||||||:|||||||||||
Db  1651 VRRADRGSVPIVNLKDEVLSPSWDSLFSGSGQLQPGARIFSFDGRDVLRHPAWPQKSVW 1710

Qy   121 HGSDPSGRRLTDSYCETWRTEAPATGQASSLLAGRLLEQBAASCRHAFVVLCIENSVMT  180
           |||||||||||:|||||||||||||||||||||||||||||||| ||| :::||||||||
Db  1711 HGSDPSGRRLMESYCETWRTETTGATGQASSLLSGRLLEQKAASCHNSYIVLCIENSFMT 1770

Qy   181 SFSK  184
           ||||
Db  1771 SFSK 1774

RESULT 3
ID   CA1E_HUMAN      STANDARD;      PRT;   1388 AA.
AC   P39059;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Collagen alpha 1(XV) chain precursor.
GN   COL15A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Umbilical cord;
RX   MEDLINE=94148920; PubMed=8106446;
RA   Kivirikko S., Heinamaki P., Rehn M.V., Horkanen N., Myers J.C.,
RA   Pihlajaniemi T.;
RT   "Primary structure of the alpha 1 chain of human type XV collagen and
RT   exon-intron organization in the 3' region of the corresponding
RT   gene.";
RL   J. Biol. Chem. 269:4773-4779(1994).
RN   [2]
RP   SEQUENCE OF 1-569 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=94140817; PubMed=8307960;
RA   Muragaki Y., Abe N., Ninomiya Y., Olsen B.R., Ooshima A.;
RT   "The human alpha 1(XV) collagen chain contains a large amino-terminal
RT   non-triple helical domain with a tandem repeat structure and homology
RT   to alpha 1(XVIII) collagen.";
RL   J. Biol. Chem. 269:4042-4046(1994).
RN   [3]
RP   SEQUENCE OF 544-1252 FROM N.A.
RX   MEDLINE=93066196; PubMed=1279671;
RA   Myers J.C., Kivirikko S., Gordon M.K., Dion A.S., Pihlajaniemi T.;
RT   "Identification of a previously unknown human collagen chain, alpha
RT   1(XV), characterized by extensive interruptions in the triple-helical
RT   region.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:10144-10148(1992).
CC   -!- TISSUE SPECIFICITY: Expressed predominantly in internal organs
CC       such as adrenal gland, pancreas and kidney.
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC       unit (G-X-Y) are hydroxylated in some or all of the chains.
CC   -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC       INTERRUPTED HELICES (FACIT) FAMILY.
CC   -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
DR   EMBL; U34609; AAC52903.1; JOINED.
DR   EMBL; U34610; AAC52903.1; JOINED.
DR   EMBL; U34611; AAC52903.1; JOINED.
DR   EMBL; U34612; AAC52903.1; JOINED.
DR   EMBL; U34613; AAC52903.1; JOINED.
DR   EMBL; U11636; AAC52178.1; -.
DR   EMBL; U11637; AAC52179.1; -.
DR   EMBL; L22545; AAA19787.1; -.
DR   EMBL; AF257775; AAF69009.1; -.
DR   PIR; A56101; A56101.
DR   PDB; 1KOE; 16-FEB-99.
DR   PDB; 1DY0; 11-APR-00.
DR   PDB; 1DY1; 21-JAN-01.
DR   MGD; MGI:88451; Col18a1.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   GO; GO:0001525; P:angiogenesis; IMP.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 8.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
DR   PROSITE; PS50038; FZ; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
KW   Alternative promoter usage; 3D-structure.
FT   SIGNAL       1     26       POTENTIAL.
FT   CHAIN       27   1774       COLLAGEN ALPHA 1(XVIII) CHAIN.
FT   CHAIN     1591   1774       ENDOSTATIN.
FT   DOMAIN     245    433       TSP N-TERMINAL.
FT   DOMAIN     365    482       FZ.
FT   DOMAIN      27    785       NONHELICAL REGION 1 (NC1).
FT   DOMAIN     786    812       TRIPLE-HELICAL REGION 1 (COL1).
FT   DOMAIN     813    822       NONHELICAL REGION 2 (NC2).
FT   DOMAIN     823    896       TRIPLE-HELICAL REGION 2 (COL2).
FT   DOMAIN     897    920       NONHELICAL REGION 3 (NC3).
FT   DOMAIN     921   1042       TRIPLE-HELICAL REGION 3 (COL3).
FT   DOMAIN    1043   1065       NONHELICAL REGION 4 (NC4).
FT   DOMAIN    1066   1148       TRIPLE-HELICAL REGION 4 (COL4).
FT   DOMAIN    1149   1162       NONHELICAL REGION 5 (NC5).
FT   DOMAIN    1163   1204       TRIPLE-HELICAL REGION 5 (COL5).
FT   DOMAIN    1205   1217       NONHELICAL REGION 6 (NC6).
FT   DOMAIN    1218   1290       TRIPLE-HELICAL REGION 6 (COL6).
FT   DOMAIN    1291   1300       NONHELICAL REGION 7 (NC7).
FT   DOMAIN    1301   1333       TRIPLE-HELICAL REGION 7 (COL7).
FT   DOMAIN    1334   1345       NONHELICAL REGION 8 (NC8).
FT   DOMAIN    1346   1369       TRIPLE-HELICAL REGION 8 (COL8).
FT   DOMAIN    1370   1376       NONHELICAL REGION 9 (NC9).
FT   DOMAIN    1377   1428       TRIPLE-HELICAL REGION 9 (COL9).
FT   DOMAIN    1429   1441       NONHELICAL REGION 10 (NC10).
FT   DOMAIN    1442   1459       TRIPLE-HELICAL REGION 10 (COL10).
FT   DOMAIN    1460   1774       NONHELICAL REGION 11 (NC11).
FT   CARBOHYD   354    354       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD   361    361       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD   585    585       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD   947    947       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   DISULFID  1623   1763
FT   DISULFID  1725   1755
FT   SITE      1351   1353       CELL ATTACHMENT SITE (POTENTIAL).
FT   VARSPLIC     1    459       Missing (in isoform 3).
FT                               /FTId=VSP_001159.
FT   VARSPLIC   460    486       AGDRLPVVCASLPSQEDGCVFIGPAA -> MAPRWHLLDV
FT                               LTESLVLLVARVSWAEP (in isoform 3).
FT                               /FTId=VSP_001158.
FT   VARSPLIC   240    486       Missing (in isoform 2).
FT                               /FTId=VSP_008303.
FT   CONFLICT  1147   1147       P -> L (IN REF. 4).
FT   CONFLICT  1194   1194       P -> F (IN REF. 4).
FT   CONFLICT  1211   1211       A -> R (IN REF. 4).
```

```
Db    1274 PIVNLKGGVLFNWDSIFSGHGGQFNNHIPIYSFDGRDIMTDPSWQKVIWHGSSPHGVR 1333
           :||||||||||||||||||||||||||| ||||||||||||:||||:|:|||||
Qy     130 LTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 184
           |:|||| ||||||||| ||:|:|: |:||||:|||||||||||||||||||||||
Db    1334 LVDNYCEAWRTADTAVTGLASPLSTGKILDQKAYSCANRLIVLCIENSFMTDARK 1388
```

RESULT 4

```
ALG8_PSRAE
ID   ALG8_PSRAE     STANDARD;      PRT;    494 AA.
AC   Q52463; Q9HY70;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Glycosyl transferase alg8 (EC 2.4.1.-).
GN   ALG8 OR PA3541.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=8830;
RX   MEDLINE=94124011; PubMed=8294014;
RA   Maharaj R., May T.B., Wang S.-K., Chakrabarty A.M.;
RT   "Sequence of the alg8 and alg44 genes involved in the synthesis of
RT   alginate by Pseudomonas aeruginosa.";
RL   Gene 136:267-269(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- FUNCTION: Possibly a processive enzyme that polymerizes GDP-
CC       mannuronic acid.
CC   -!- PATHWAY: Alginate biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 2.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L22611; AAC36875.2; ALT_INIT.
DR   EMBL; AE004774; AAG06929.1; -.
DR   PIR; A83204; A83204.
KW   Alginate biosynthesis; Transferase; Glycosyltransferase;
KW   Transmembrane; Complete proteome.
FT   TRANSMEM     12     34       POTENTIAL.
FT   TRANSMEM     49     71       POTENTIAL.
FT   TRANSMEM    381    403       POTENTIAL.
FT   TRANSMEM    423    445       POTENTIAL.
FT   CONFLICT    115    115       A -> R (IN REF. 1).
SQ   SEQUENCE   494 AA;  56456 MW;  D89A5627B913FDF5 CRC64;

Query Match             9.0%;  Score 87;  DB 1;  Length 494;
Best Local Similarity  24.6%;  Pred. No. 0.8;
Matches   45;  Conservative   20;  Mismatches   68;  Indels   50;  Gaps    7;

Qy     4 QDFQLVLHLVALNSPQPGGMRGIRGADF----------QCFQQARAAGLAGTFRAFLSSR 53
         ||| |: ::: |||  ::: |:| |:||          :|| |:|:|| ||||  ||:
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L25286; AAA58429.1; -.
DR   EMBL; D21230; BAA04762.1; -.
DR   EMBL; L01697; -; NOT_ANNOTATED_CDS.
DR   PIR; A53317; A53317.
DR   HSSP; P39061; 1KOB.
DR   Genew; HGNC:2192; COL15A1.
DR   MIM; 120325; -.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008995; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 5.
DR   Pfam; PF02210; TSPN; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cell adhesion; Collagen; Glycoprotein; Signal.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   1388       COLLAGEN ALPHA 1(XV) CHAIN.
FT   DOMAIN       40    228       TSP N-TERMINAL.
FT   DOMAIN      229    555       NONHELICAL REGION 1 (NC1).
FT   DOMAIN      556    573       TRIPLE-HELICAL REGION 1 (COL1).
FT   DOMAIN      574    618       NONHELICAL REGION 2 (NC2).
FT   DOMAIN      619    732       TRIPLE-HELICAL REGION 2 (COL2).
FT   DOMAIN      733    763       NONHELICAL REGION 3 (NC3).
FT   DOMAIN      764    798       TRIPLE-HELICAL REGION 3 (COL3).
FT   DOMAIN      799    822       NONHELICAL REGION 4 (NC4).
FT   DOMAIN      823    867       TRIPLE-HELICAL REGION 4 (COL4).
FT   DOMAIN      868    878       NONHELICAL REGION 5 (NC5).
FT   DOMAIN      879    949       TRIPLE-HELICAL REGION 5 (COL5).
FT   DOMAIN      950    983       NONHELICAL REGION 6 (NC6).
FT   DOMAIN      984   1013       TRIPLE-HELICAL REGION 6 (COL6).
FT   DOMAIN     1014   1027       NONHELICAL REGION 7 (NC7).
FT   DOMAIN     1028   1045       TRIPLE-HELICAL REGION 7 (COL7).
FT   DOMAIN     1046   1052       NONHELICAL REGION 8 (NC8).
FT   DOMAIN     1053   1107       TRIPLE-HELICAL REGION 8 (COL8).
FT   DOMAIN     1108   1117       NONHELICAL REGION 9 (NC9).
FT   DOMAIN     1118   1132       TRIPLE-HELICAL REGION 9 (COL9).
FT   DOMAIN     1133   1388       NONHELICAL REGION 10 (NC10).
FT   REPEAT      358    555       4 X TANDEM REPEATS.
FT   REPEAT      409    459       1.
FT   REPEAT      460    509       2.
FT   REPEAT      510    555       3.
FT   REPEAT                       4.
FT   CARBOHYD    306    306       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    324    324       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    687    687       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    807    807       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    814    814       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD   1046   1046       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CONFLICT     10     10       C -> S (IN REF. 2).
FT   CONFLICT     49     49       D -> V (IN REF. 2).
FT   CONFLICT     95     95       L -> A (IN REF. 2).
FT   CONFLICT    150    150       P -> V (IN REF. 2).
FT   CONFLICT    204    204       M -> V (IN REF. 2).
FT   CONFLICT    409    409       R -> A (IN REF. 2).
SQ   SEQUENCE  1388 AA;  141930 MW;  60822AD925A3093D CRC64;

Query Match            54.5%;  Score 526;  DB 1;  Length 1388;
Best Local Similarity  56.6%;  Pred. No. 1.2e-42;
Matches   99;  Conservative   27;  Mismatches   45;  Indels    4;  Gaps    1;

Qy    10 LHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGV 69
         |:||: |||| :|||||||||: ||||||||||||||||||:||:|:||::|||| |:|
Db  1218 LHLAALNMPFSGDIR-----ADFQCFKQARAAGLLSTYRAFLSSHLQDLSTIVRKAERYSL 1273

Qy    70 PVVNLRDEVLFSWBEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRR 129
         |:||||| |||: :||| | || : ::|||||||||| ||||  | |||:|||| | |
```

```
Db    39 KDFILLIGAVGIWRYSMGGVHFLRGMLFLHVVTPYRRRVRQLGSAADPSHVFLMVTSFR  98

Qy    54 LQDL-----YSIVRRADRTGVP------VVNLRDEVLFPS-WEAL----------       86
         :::       :||||||| ||       |||||||||| :|:|
Db    99 IDALTTAMVRSVIREAIDSGYPTTVVCSIVEMSDEVLVRSLWEKMNPPDRVSLDFVRIP  158

Qy    87 FSGSEGQLKPGARIFS----------FDGRDVLQHPAWPRKSVWHGSDPS-GRRLTDS  133
         ||||||||||||||||          ||||||||||||||||||||||   :|| ||
Db   159 GTGKRDGLAYGFRAISRHLPDDDAVVAVIDGDTVLDHGVVKKTVPWFKLFPNVGGLTTNE 218

Qy   134 YCE 136
         :|:
Db   219 FCE 221


RESULT 5
GLI_CHICK
ID   GLI_CHICK      STANDARD;      PRT;   556 AA.
AC   P55878;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein GLI1 (GLI) (Fragment).
GN   GLI1 OR GLI.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97105842; PubMed=8948590;
RA   Marigo V., Johnson R.L., Vortkamp A., Tabin C.J.;
RT   "Sonic hedgehog differentially regulates expression of GLI and GLI3
RT   during limb development.";
RL   Dev. Biol. 180:273-283(1996).
CC   -!- FUNCTION: MAY REGULATE THE TRANSCRIPTION OF SPECIFIC GENES DURING
CC       NORMAL DEVELOPMENT. MAY PLAY A ROLE IN CRANIOFACIAL DEVELOPMENT
CC       AND DIGITAL DEVELOPMENT, AS WELL AS DEVELOPMENT OF THE CENTRAL
CC       NERVOUS SYSTEM AND GASTROINTESTINAL TRACT. IMPLICATED IN THE
CC       TRANSDUCTION OF SHH SIGNAL (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
```
```
DR   EMBL; U60762; AAB51659.1; -.
DR   HSSP; P08151; 2GLI.
DR   InterPro; IPR007087; ZnF_C2H2.
DR   Pfam; PF00096; zf-C2H2; 5.
DR   SMART; SM00355; ZnF_C2H2; 5.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 5.
KW   Zinc-finger; Metal-binding; DNA-binding; Transcription regulation;
KW   Nuclear protein; Repeat.
FT   ZN_FING     247    272       C2H2-TYPE.
FT   ZN_FING     280    307       C2H2-TYPE.
FT   ZN_FING     313    337       C2H2-TYPE.
FT   ZN_FING     343    368       C2H2-TYPE.
FT   ZN_FING     374    399       C2H2-TYPE.
FT   NON_TER     556    556
SQ   SEQUENCE   556 AA;   60215 MW;  722D2AA5A1CA4D98 CRC64;

Query Match             8.9%;  Score 85.5;  DB 1;  Length 556;
Best Local Similarity   26.4%;  Pred. No. 1.3;
Matches   53;  Conservative   20;  Mismatches   75;  Indels   53;  Gaps   10;
```

```
Qy     2 THQDFQLVLHLVALNSPQPGGMRGIRGADFQCEQQARAAGLAGTFRAF----LSSRLQDL  57
         :  |:                                             ||||||||
Db    47 SHHGYGLV-----PGTEHPGG----AADGSRFSTPRGAGKLGKKRALSISPLSDSSVDL   96

Qy    58 YSIVRRADRTGPVPVVNLRDEVLFPSWEAL------------FSGSEGQLKPGARIFS-  102
         :::|                                           :::
Db    97 QTVIRTSPNSLVAFINSRCASAGGSYGHLSISTISPSLGYQNPPGQQKGQGQLFSHTPPL 156

Qy   103 -------FDGRDVLQHPAWPRKSVWHGSDPSGRR-----LTDSYCETWRTE---APAA  145
         ::|                                      :| ::  ::  :
Db   157 PPCCSHETLSSRPGLLHFTPARGTIKHCQQLKLERSLSSPLTAKYPEE-KSEGDISSPAS 215

Qy   146 TGQASSLLAGRL-----LEQE  161
         ||   |  :|
Db   216 TGTQDPLL-GMLSVRDDLEKE  235


RESULT 6
SR13_HUMAN
ID   SR13_HUMAN     STANDARD;      PRT;   995 AA.
AC   Q9Y3M8;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   StAR-related lipid transfer protein 13 (StARD13) (START domain-
DE   containing protein 13) (46H23.2).
GN   STARD13 OR GT650.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rhodes S.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Hunt A.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: May function as a GTPase-activating protein.
CC   -!- SIMILARITY: Contains 1 Rho-GAP domain.
CC   -!- SIMILARITY: Contains 1 START domain.
```
```
DR   EMBL; AL049801; CAB42562.1; -.
DR   EMBL; Z84483; CAC94774.1; -.
DR   PIR; H59432; H59432.
DR   Genew; HGNC:19164; STARD13.
DR   InterPro; IPR008936; Rho_GAP.
DR   InterPro; IPR000198; RhoGAP.
DR   InterPro; IPR002913; START.
DR   Pfam; PF00620; RhoGAP; 1.
DR   Pfam; PF01852; START; 1.
DR   SMART; SM00324; RhoGAP; 1.
DR   SMART; SM00234; START; 1.
DR   PROSITE; PS50238; RHOGAP; 1.
DR   PROSITE; PS50848; START; 1.
KW   GTPase activation.
FT   DOMAIN     545    750       RHO-GAP.
FT   DOMAIN     781    989       START.
SQ   SEQUENCE   995 AA;  111191 MW;  3F6O8FA94A4EF8BF CRC64;

Query Match             8.5%;  Score 82;  DB 1;  Length 995;
Best Local Similarity   23.0%;  Pred. No. 5.4;
Matches   37;  Conservative   26;  Mismatches   52;  Indels   46;  Gaps   7;
```

```
Qy    52  SRLQDLYSIVRRADRTGVP-------------------VVNLRDEVLFPSWEALFSGSEGQLKPG  97
              |||:||||::||                         :|:|         |||||||||||||:||:||
Db    36  SRVDDLYTLLPRGDRNGSPGGTGMRNTTSSESVLTDLSEPEVCSIHSESSGGSDRSQPG  95

Qy    98  ---------ARIFSFD---GRDVLQHPAWPRKSVWHGSDPSGRRLTDSY---CETWR  139
                      :|:         ||||| || |         :|        :|         |:|:|
Db    96  QCCTDNPVMLDAPLVSSSLPQPPRDVLNHPFHFK------NEKPTRARAKSFLKRMETLR  149

Qy   140  TEAPAATGQASS-------LLAGRILLEQEAASCRHAFVVLICIE  175
              :                     |:      | |||      |  ||| |
Db   150  GKGAHGRHKGSGRTGGLVISGPMLQQBPESFK---AMQCIQ  187

RESULT 7
NME3_HUMAN
ID   NME3_HUMAN            STANDARD;        PRT;   1233 AA.
AC   Q14957;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Glutamate [NMDA] receptor subunit epsilon 3 precursor (N-methyl
DE   D-aspartate receptor subtype 2C) (NR2C) (NMDAR2C).
GN   GRIN2C.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97189248; PubMed=9037519;
RA   Lin Y.J., Bovetto S., Carver J.M., Giordano T.;
RT   "Cloning of the cDNA for the human NMDA receptor NR2C subunit and its
RT   expression in the central nervous system and periphery.";
RL   Brain Res. Mol. Brain Res. 43:57-64(1996).
CC   -!- FUNCTION: NMDA receptor subtype of glutamate-gated ion channels
CC       with high calcium permeability and voltage-dependent sensitivity
CC       to magnesium. Mediated by glycine.
CC   -!- SUBUNIT: Heterodimer of an epsilon subunit and a zeta subunit.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Mainly in brain with predominant expression is
CC       in the cerebellum, also present in the hippocampus, amygdala,
CC       caudate nucleus, corpus callosum, subthalamic nuclei and thalamus.
CC       Detected in the heart, skeletal muscle and pancreas.
CC   -!- SIMILARITY: Belongs to the ligand-gated ionic channel family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L76224; AAB88096.1; -.
DR   HSSP; P19491; 1GR2.
DR   Genew; HGNC:4587; GRIN2C.
DR   MIM; 138254; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0004972; F:N-methyl-D-aspartate selective glutamate re. . .; TAS.
DR   GO; GO:0007215; P:glutamate signaling pathway; TAS.
DR   GO; GO:0006810; P:transport; TAS.
DR   InterPro; IPR001828; ANF_receptor.
DR   InterPro; IPR001320; Ion_glu_receptor.
DR   InterPro; IPR001508; NMDA_receptor.
DR   InterPro; IPR001311; SBP/glu_receptor.
DR   Pfam; PF01094; ANF_receptor; 1.
DR   Pfam; PF00060; lig_chan; 1.
DR   PRINTS; PR00177; NMDARECEPTOR.
DR   SMART; SM00079; PBPe; 1.
KW   Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;
KW   Ionic channel; Magnesium.
```

```
FT   SIGNAL          1     19                              POTENTIAL.
FT   CHAIN          20   1233                              GLUTAMATE [NMDA] RECEPTOR SUBUNIT
FT                                                         EPSILON 3.
FT   TRANSMEM      554    574                              1 (POTENTIAL).
FT   TRANSMEM      597    617                              2 (POTENTIAL).
FT   TRANSMEM      627    647                              3 (POTENTIAL).
FT   TRANSMEM      815    835                              4 (POTENTIAL).
FT   SITE          612    612                              FUNCTIONAL DETERMINANT OF NMDA
FT                                                         RECEPTORS (BY SIMILARITY).
FT   CARBOHYD       70     70                              N-LINKED (GLCNAC. . .)  (POTENTIAL).
FT   CARBOHYD      337    337                              N-LINKED (GLCNAC. . .)  (POTENTIAL).
FT   CARBOHYD      438    438                              N-LINKED (GLCNAC. . .)  (POTENTIAL).
FT   CARBOHYD      539    539                              N-LINKED (GLCNAC. . .)  (POTENTIAL).
SQ   SEQUENCE   1233 AA;  134239 MW;  700F24629ED0D699 CRC64;

Query Match             8.5%;  Score 82;  DB 1;  Length 1233;
Best Local Similarity   32.8%;  Pred. No. 6.9;
Matches  41;  Conservative   9;  Mismatches  49;  Indels  26;  Gaps   7;

Qy    58  YSIVRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRK  117
              ||||:|||||         |||:|||       :||: :|||      |             |  |:
Db  1029  YSSFPRADRSGRPFL------PLFPELEDLPLLGKEQL---AR------REALLHAAWAR-  1073

Qy   118  SVWHGSDPSGRRLTDSYCETWR--TEAPA-ATGQASSLLAG----RLLEQEAASCRHAFV  170
              |||       :| | |                :|           :|          ||||
Db  1074  ---GSRPRHASLPSSVAEAFARPSSLPAGCTGPACARPDGHSACRRLAQAQSMCLPIYR  1129

Qy   171  VLCIE  175
Db  1130  EACQE  1134
```

```
RESULT 8
NRT1_CHICK
ID   NRT1_CHICK            STANDARD;        PRT;    312 AA.
AC   P55806;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   NAD(P)(+)--arginine ADP-ribosyltransferase 1 precursor (EC 2.4.2.31)
DE   (Mono(ADP-ribosyl)transferase 1) (AT1).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=White leghorn; TISSUE=Bone marrow;
RX   MEDLINE=95050487; PubMed=7961658;
RA   Tsuchiya M., Hara N., Yamada K., Osago H., Shimoyama M.;
RT   "Cloning and expression of cDNA for arginine-specific ADP-
RT   ribosyltransferase from chicken bone marrow cells.";
RL   J. Biol. Chem. 269:27451-27457(1994).
CC   -!- CATALYTIC ACTIVITY: NAD(+) + L-arginine = nicotinamide + N(2)-
CC       (ADP-D-ribosyl)-L-arginine.
CC   -!- SUBCELLULAR LOCATION: THE MATURE ENZYME IS PROBABLY SECRETED
CC       FROM GRANULOCYTES INTO THE EXTRACELLULAR SPACE.
CC   -!- SIMILARITY: Belongs to the Arg-specific ADP-ribosyltransferase
CC       family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D31864; BAA06664.1; -.
DR   PIR; A55461; A55461.
DR   InterPro; IPR000768; ART.
DR   Pfam; PF01129; ART; 1.
```

```
DR   PRINTS; PR00970; RIBTRNSFRASE.
DR   PROSITE; PS01291; ART; 1.
KW   Transferase; Glycosyltransferase; NAD; Signal; Zymogen.
FT   SIGNAL        1     20       POTENTIAL.
FT   PROPEP       21     31       NAD(P)(+)--ARGININE ADP-
FT   CHAIN        32    266       RIBOSYLTRANSFERASE 1.
                                  POTENTIAL.
FT   PROPEP      267    312       BY SIMILARITY.
FT   ACT_SITE    231    231
SQ   SEQUENCE    312 AA;  35318 MW;  B82980439BC904FC CRC64;

Query Match              8.4%;  Score 81;  DB 1;  Length 312;
Best Local Similarity   26.3%;  Pred. No. 1.8;
Matches   50;  Conservative 13;  Mismatches  51;  Indels 76;  Gaps 12;

Qy     2  THQD----------------FQLVLHLVALNSPQP------GGMRGIRGADFQCFQQARAAGLAG 44
           :||                 :|||                    |||:|||
Db   128  SHQDYIHSYHFKTLHFFLTQALFALRASQPRCYVVRGVRGIR------FMTQR--GKSV 179

Qy    45  TFRAFLSSRL---------QDLYSIVRRADRTGVPVVNL----RDEVLFPSWRAL----- 86
           :||:                :|:          |||:          :||||
Db   180  RFGQFTSTSLRKEATVNFGQDTLFVVKTC--YGVPIKQFSFFPSEDDVLIPPFEVFEVIN 237

Qy    87  FSGSSGQ-----------------LKPGARIFSFDGRDV-------LQHPAWP-RKSVW 120
           |||:||                   ||||             ||          ||||:|
Db   238  FSNDRGSVKIQLHSKGKMSTHNCELLKPQGGQWGRGHQEVGLGLSPGLSLPVLPCRRRVW 297

Qy   121  HG-----SDP 125
           ||      |
Db   298  EGLGHREGDP 307


RESULT 9
P47K_PSECL
ID   P47K_PSECL    STANDARD;      PRT;    419 AA.
AC   P31521;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   47 kDa protein (P47K).
OS   Pseudomonas chlororaphis (Pseudomonas aureofaciens).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=333;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B23;
RX   MEDLINE=91193202; PubMed=2013568;
RA   Nishiyama M., Horinouchi S., Kobayashi M., Nagasawa T., Yamada H.,
RA   Beppu T.;
RT   "Cloning and characterization of genes responsible for metabolism of
RT   nitrile compounds from Pseudomonas chlororaphis B23.";
RL   J. Bacteriol. 173:2465-2472(1991).
CC   -!- FUNCTION: THE PRESENCE OF P47K IS CRITICAL FOR THE EXPRESSION
CC       OF THE NITRILE HYDRATASE GENES. MAY STABILIZE OR ACTIVATE THE
CC       NITRILE HYDRATASE PROTEINS.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; D90216; BAA14247.1; -.
DR   PIR; D42725; D42725.
DR   InterPro; IPR003495; CobW.
DR   Pfam; PF02492; CobW; 1.
SQ   SEQUENCE    419 AA;  46666 MW;  FF5113800E27FF0C CRC64;

Query Match              8.0%;  Score 77;  DB 1;  Length 419;
Best Local Similarity   24.0%;  Pred. No. 6.1;
```

```
Matches   43;  Conservative  26;  Mismatches  26;  Conservative  76;  Indels  34;  Gaps   9;

Qy    26  IRGADFQCFQQARAAGLAGTFRAFLSSR-LQDLYSIVRRADRTGVPVVNLRDEVLFPSWE 84
           :|:    ||  :|          :|:||||   |:|:  |||| ||    ||||:|||: ::
Db   141  VDGSQFQALLESTDTVARADTEAHTSTRHADL--LIEQVEYANVILVNKRDLIDEPGYQ 198

Qy    85  ALPSGSEGQLKPGARIFS--------------------FDGRDVLQHPAWPRKSVWHGSDPSGR 128
           |: :|:|  |||||||                       ::|
Db   199  AVHAILAG-LNPSARIMPVAHGNVALSSLLDTHLFDLPSLAASPGWMRK--MEATDTPAS 255

Qy   129  RLTDSYCET--WRTEAPAATGQASSLLA----GRLLEQEA--ASCRHAFVVLCIEN 176
           |  :| :|       :||   ||:|||         ||:||||   :: |   :|
Db   256  E-SDTYGVTSWVYRERAPFHPQRLLEFLQKPWHNGRLLRSKGYFWLASRHLEIGLLAQS 313


RESULT 10
CAR4_MOUSE
ID   CAR4_MOUSE    STANDARD;      PRT;    953 AA.
AC   Q8BHB0; Q8BUT6;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Caspase recruitment domain protein 4.
GN   CARD4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NOD; TISSUE=Cerebellum, and Spleen;
RX   MEDLINE=22354683; PubMed=12466851;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasaki Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Magiott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
RN   [2]
RP   SEQUENCE FROM N.A., AND VARIANT ALA-884.
RC   STRAIN=Czech II, and FVB/N; TISSUE=Breast cancer;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
```

```
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Enhances caspase-9-mediated apoptosis. Induces NF-kappa-
CC       B activity via RICK (CARDIAK, RIP2) and IKK-gamma. Confers
CC       responsivenes to intracellular bacterial lipopolysaccharides
CC       (LPS) (By similarity)
CC   -!- SUBUNIT: Self-associates. Binds to caspase-9 and RICK by CARD-CARD
CC       interaction (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Contains 1 CARD domain.
CC   -!- SIMILARITY: Contains 1 NACHT domain.
CC   -!- SIMILARITY: Contains 9 leucine-rich (LRR) repeats.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AK082663; BAC38566.1; -.
DR   EMBL; AK089562; BAC40940.1; -.
DR   EMBL; BC043670; AAH42670.1; -.
DR   EMBL; BC043670; AAH43670.1; -.
DR   MGD; MGI:1341839; Card4.
DR   InterPro; IPR001315; CARD.
DR   InterPro; IPR007091; LRR_RNinh.
DR   InterPro; IPR007111; NACHT_NTPase.
DR   Pfam; PF00619; CARD; 1.
DR   Pfam; PF05729; NACHT; 1.
DR   PROSITE; PS50209; CARD; 1.
DR   PROSITE; PS50837; NACHT; 1.
KW   Apoptosis; ATP-binding; Repeat; Leucine-rich repeat; Polymorphism.
FT   DOMAIN     15    107       CARD.
FT   DOMAIN    196    531       NACHT.
FT   NP_BIND   202    209       ATP (POTENTIAL).
FT   REPEAT     17     42       LRR 1.
FT   REPEAT    702    725       LRR 2.
FT   REPEAT    727    750       LRR 3.
FT   REPEAT    755    778       LRR 4.
FT   REPEAT    783    806       LRR 5.
FT   REPEAT    839    862       LRR 6.
FT   REPEAT    867    890       LRR 7.
FT   REPEAT    895    918       LRR 8.
FT   REPEAT    923    946       LRR 9.
FT   VARIANT   884    884       S -> A (in strain Czech II).
SQ   SEQUENCE   953 AA;  107739 MW;  39C6396621CEB1A58 CRC64;

Query Match             7.9%;  Score 76.5;  DB 1;  Length 953;
Best Local Similarity   24.4%;
Matches   47;  Conservative   28;  Mismatches   57;  Indels   61;  Gaps   11;

Qy    2  THQDFQLVLHIVALNSPQPGGMRGI------RGADFQCFQQARAAGLAGTFRAFLSSRL  54
                                          : | || | |        |:  ||| | |
Db  158  TIYMD--TIMELVGFNNENLGSLGGLDCLLDHSTGVLNEHGETVFVFGDAGVGKSMLQRL 215

Qy   55  QDL------------YSIVRRADRTGVPVVNLRDEVLFPSW------EALF  87
              :              :| :| |      :| ||: | :| | :|
Db  216  QSLWASGRLTSTAKFFHFRCRMFSCFKESD-----MLSLQD-LLFXHFCYPEQDPEEVF 269

Qy   88  SGSEGQLK-PGARIFSPDGRDVLQHPAWPRKSVWHGSPDSGSGRRLTDSYCETWRTEAPAAT 146
           | :: |||| |              ||  |:  | | | |  |
```

```
Db  270  S---FLLRFPHTALFTPGLDELH----------SDFDLSRVPDSCC-PWEPAHPLVL 313
              :| || | | |                          ||:|||:|||
Qy  147  GQASSLLAGRLLE 159
              :: ||||| |||
Db  314  --LANLLSGRLLK 324

RESULT 11
R1AB_CVHSA
ID   R1AB_CVHSA     STANDARD;      PRT;   7073 AA.
AC   P59641; Q808C0; Q80BV7; Q80BV8; Q80BS1;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Replicase polyprotein 1ab (pp1ab) (ORF1AB) [Includes: Replicase
DE   polyprotein 1a (pp1a) (ORF1A) [Contains: Leader protein; p65 homolog;
DE   NSP1 (EC 3.4.24.-) (Papain-like proteinase) (PL-PRO); 3C-like
DE   proteinase (EC 3.4.24.-) (3CLp) (NSP2); NSP3 (3CL-PRO) (NSP2); NSP4; NSP5;
DE   NSP6; Growth factor-like peptide (GFL) (NSP7); RNA-directed RNA
DE   polymerase (EC 2.7.7.48) (RdRp) (Pol) (NSP9); Helicase (Hel) (NSP10);
DE   NSP11; NSP12; NSP13].
OS   Human coronavirus (strain SARS) (HCoV-SARS) (SARS-CoV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=227859;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=isolate Urbani;
RX   MEDLINE=22660724; PubMed=12730500;
RA   Rota P.A., Oberste M.S., Monroe S.S., Nix W.A., Campagnoli R.,
RA   Icenogle J.P., Penaranda S., Bankamp B., Maher K., Chen M.-H.,
RA   Tong S., Tamin A., Lowe L., Frace M., DeRisi J.L., Chen Q., Wang D.,
RA   Erdman D.D., Peret T.C.T., Burns C., Ksiazek T.G., Rollin P.E.,
RA   Sanchez A., Liffick S., Holloway B., Limor J., McCaustland K.,
RA   Olsen-Rasmussen M., Fouchier R., Guenther S., Osterhaus A.D.M.E.,
RA   Drosten C., Pallansch M.A., Anderson L.J., Bellini W.J.;
RT   "Characterization of a novel coronavirus associated with severe acute
RT   respiratory syndrome.";
RL   Science 300:1394-1399(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=isolate Tor2;
RX   MEDLINE=22660725; PubMed=12730501;
RA   Marra M.A., Jones S.J.M., Astell C.R., Holt R.A., Brooks-Wilson A.,
RA   Butterfield Y.S.N., Khattra J., Asano J.K., Barber S.A., Chan S.Y.,
RA   Cloutier J., Coughlin S.M., Freeman D., Girn N., Griffith O.L.,
RA   Leach S.R., Mayo M., McDonald H., Montgomery S.B., Pandoh P.K.,
RA   Petrescu A.S., Robertson A.G., Schein J.E., Siddiqui A., Smailus D.E.,
RA   Stott J.M., Yang G.S., Plummer F., Andonov A., Artsob H., Bastien N.,
RA   Bernard K., Booth T.F., Bowness D., Czub M., Drebot M., Fernando L.,
RA   Flick R., Garbutt M., Gray M., Grolla A., Jones S., Feldmann H.,
RA   Meyers A., Kabani A., Li Y., Normand S., Stroher U., Tipples G.A.,
RA   Tyler S., Vogrig R., Ward D., Watson B., Brunham R.C., Krajden M.,
RA   Petric M., Skowronski D.M., Upton C., Roper R.L.;
RT   "The genome sequence of the SARS-associated coronavirus.";
RL   Science 300:1399-1404(2003).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=isolate CUHK-Su10, and Isolate CUHK-W1;
RX   MEDLINE=22737955; PubMed=12853594;
RA   Tsui S.K.W., Chim S.S.C., Lo Y.M.D.;
RT   "Coronavirus genomic-sequence variations and the epidemiology of the
RT   severe acute respiratory syndrome.";
RL   New Engl. J. Med. 349:187-188(2003).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=isolate GZ50, Isolate SZ3, and Isolate SZ16;
RX   MEDLINE=22911660; PubMed=12958166;
RA   Guan Y., Zheng B.J., He Y.Q., Liu X.L., Zhuang Z.X., Cheung C.L.,
RA   Luo S.W., Li P.H., Zhang L.J., Guan Y.J., Butt K.M., Wong K.L.,
RA   Chan K.W., Lim W., Shortridge K.F., Yuen K.Y., Peiris J.S.M.,
RA   Poon L.L.M.;
RT   "Isolation and characterization of viruses related to the SARS
```

RT   coronavirus from animals in southern China.";
RL   Science 302:276-278(2003).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Isolate HKU-39849;
RX   MEDLINE=22758472; PubMed=12876307;
RA   Zeng F.Y., Chan C.W., Chan M.N., Chen J.D., Chow K.Y.C., Hon C.C.C.,
RA   Hui R.K.H., Li J., Li V.Y.Y., Wang C.Y., Yuen K.Y., Peiris J.S.M., Guan Y., Zheng B.,
RA   Poon L.L.M., Chan K.H., Yuen K.Y., Peiris J.S.M., Leung F.C.;
RT   "The complete genome sequence of severe acute respiratory syndrome
RT   coronavirus strain HKU-39849 (HK-39).";
RL   Exp. Biol. Med. 228:866-873(2003).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Isolate Sin2500, Isolate Sin2677, Isolate Sin2679,
RC   Isolate sin2774, and Isolate Sin2748;
RX   MEDLINE=22667074; PubMed=12781537;
RA   Ruan Y., Wei C.L., Ling A.E., Vega V.B., Thoreau H., Se Thoe S.Y.,
RA   Chia J.-M., Ng P., Chiu K.P., Lim L., Zhang T., Chan K.P., Oon L.E.L.,
RA   Ng M.L., Leo S.Y., Ng L.F.P., Ren E.C., Stanton L.W., Long P.M.,
RA   Liu E.T.;
RT   "Comparative full-length genome sequence analysis of 14 SARS
RT   coronavirus isolates and common mutations associated with putative
RT   origins of infection.";
RL   Lancet 361:1779-1785(2003).
RN   [7]
RP   ERRATUM.
RA   Ruan Y., Wei C.L., Ling A.E., Vega V.B., Thoreau H., Se Thoe S.Y.,
RA   Chia J.-M., Ng P., Chiu K.P., Lim L., Zhang T., Chan K.P., Oon L.E.L.,
RA   Ng M.L., Leo S.Y., Ng L.F.P., Ren E.C., Stanton L.W., Long P.M.,
RA   Liu E.T.;
RL   Lancet 361:1832-1832(2003).
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Isolate BJ01, Isolate BJ02, Isolate BJ03, Isolate BJ04,
RC   and Isolate GD01;
RA   Qin E., Zhu Q., Yu M., Fan B., Chang G., Si B., Yang B., Peng W.,
RA   Jiang T., Liu B., Deng Y., Liu H., Zhang Y., Wang C., Li Y., Gan Y.,
RA   Li X., Lu F., Tan G., Yang R., Cao W.S., Wang J., Chen W., Cong L.,
RA   Deng Y., Dong W., Han Y., Hu W., Lei M., Li C., Li G., Li H.,
RA   Li S., Li W., Li W., Lin W., Liu J., Liu Z., Lu H., Ni P.,
RA   Qi Q., Sun Y., Tang L., Tong Z., Wang J., Wang X., Wu Q., Xi Y.,
RA   Xu Z., Yang L., Ye C., Ye J., Zhang B., Zhang F., Zhang J., Zhang X.,
RA   Zhou J., Yang H.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Isolate TW1;
RA   Yeh S.-H., Kao C.-L., Tsai C.-Y., Liu C.-J., Chen D.-S., Chen P.-J.;
RT   "The complete genome of SARS coronavirus clone TW1.";
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Isolate FRA;
RA   Eickmann M., Becker S., Klenk H.-D., Doerr H.W., Stadler K.,
RA   Censini S., Guidotti S., Masignani V., Scarselli M., Mora M.,
RA   Donati C., Han J., Song H., Abrignani S., Covacci A., Rappuoli R.;
RT   "SARS virus is a close relative of type II coronaviruses.";
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Isolate Frankfurt 1;
RA   Thiel V., Hertzig T., Putics A., Ivanov K.A., Schelle B., Bayer S.,
RA   Scheiner B., Weinand H., Weissbrich B., Ziebuhr J.;
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
RN   [12]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Isolate TWC;
RA   Yang J.-Y., Lin J.-H., Chiu S.-C., Wang S.-F., Lee S.C., Lin Y.-C.,
RA   Hsu C.-K., Chen H.-Y., Chang J.G., Chan P.-J., Su I.-J.;
RT   "Genomic sequence of SARS isolate from the first fatal case in
RT   Taiwan.";
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.

RN   [13]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Isolate HSR 1;
RA   Canducci F., Clementi M., Poli G., Vicenzi E.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
RN   [14]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Isolate Taiwan TC1, Isolate Taiwan TC2, and Isolate Taiwan TC3;
RA   Chang J.-G.C., Lin T.-H., Chen C.-M., Lin C.-S., Chan W.-L.,
RA   Shih M.-C.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
RN   [15]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Isolate TWH, Isolate TWJ, Isolate TWK, Isolate TWS,
RC   and Isolate TWY;
RA   Shu H.Y., Wu K.M., Tsai S.F.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
RN   [16]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Isolate AS;
RA   Balotta C., Corvasce S., Violin M., Galli M., Moroni M.,
RA   Vigevani G.M., Ruan Y.J., Salemi M.;
RL   Submitted (OCT-2003) to the EMBL/GenBank/DDBJ databases.
RN   [17]
RP   SEQUENCE OF 1-322 FROM N.A.
RC   STRAIN=Isolate ZJ-HZ01;
RA   Wang Z., Cheng S., Zhang Y.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
RN   [18]
RP   SEQUENCE OF 1-507; 1655-5170 AND 6903-7073 FROM N.A.
RC   STRAIN=Isolate Shanghai LY;
RA   Yuan Z., Zhang X., Hu Y., Lan S., Wang H., Zhou Z., Wen Y.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
RN   [19]
RP   SEQUENCE OF 497-1674 AND 5158-6911 FROM N.A.
RC   STRAIN=Isolate Shanghai QXC;
RA   Yuan Z., Zhang X., Hu Y., Lan S., Wang H., Zhou Z., Wen Y.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
RN   [20]
RP   SEQUENCE OF 4993-5127 FROM N.A.
RC   STRAIN=Isolate Vietnam;
RA   Emery S., Erdman D.D., Peret T.C.T., Ksiazek T.G.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [21]
RP   SEQUENCE OF 4993-5136 FROM N.A.
RC   STRAIN=Isolate Taiwan;
RA   Lin J.-H., Chiu S.-C., Yang J.-Y., Wang S.-F., Chen H.-Y.;
RT   "Detection of a novel human coronavirus in a severe acute respiratory
RT   syndrome patient in Taiwan.";
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [22]
RP   CHARACTERIZATION OF HELICASE.
RX   MEDLINE=22890296; PubMed=12917423;
RA   Tanner J.A., Watt R.M., Chai Y.-B., Lu L.-Y., Lin M.C., Peiris J.S.,
RA   Poon L.L.M., Kung H.-F., Huang J.-D.;
RT   "The severe acute respiratory syndrome (SARS) coronavirus
RT   NTPase/helicase belongs to a distinct class of 5' to 3' viral
RT   helicases.";
RL   J. Biol. Chem. 278:39578-39582(2003).
RN   [23]
RP   3D-STRUCTURE MODELING OF 3241-3540, AND CHARACTERIZATION.
RX   MEDLINE=22690353; PubMed=12746549;
RA   Anand K., Ziebuhr J., Wadhwani P., Mesters J.R., Hilgenfeld R.;

Query Match
Best Local Similarity   7.9%;   Score 76;   DB 1;   Length 7073;
                        22.1%;   Pred. No. 2e+02;
Matches   36;   Conservative   24;   Mismatches   73;   Indels   30;   Gaps   4;

```
Qy    1 HTHQDFQLVLHLVALNS------PQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRL   54
                         |||    |:::|||||
Db 1613 HTLDESFLGRYMSALNHTKKWKFPQVGGLTSIKWADNNCYLSSVLLALQQLEVKFNAPAL 1672
Qy   55 QDLIYSIVRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSPDGRDVLQHPAW  114
```

Db   1673 |:|:| ||| | : | |:|: | | |:| :|
            QEAYYRARAGDAANFCALIL----------AYSNKTVGELG-------DVRETMTH--- 1711

Qy   115 PRKSVWHGSDPSGRRLTDSYCETWRTRTEAPAATGQASSLLAGRL 157

Db   1712 ---LIQHANLESAKRVLNVVCKHCGQKTTLITGVEAVMYMGTL 1751

RESULT 12

PPX_ECOLI
ID   PPX_ECOLI   STANDARD;   PRT;   512 AA.
AC   P29014; P76981;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Exopolyphosphatase (EC 3.6.1.11) (Exopolyase) (Metaphosphatase).
GN   PPX OR B502 OR C3020 OR Z3765 OR ECS3364.
OS   Escherichia coli,
OS   Escherichia coli O6, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562, 217992, 83334;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC   STRAIN=K12;
RX   MEDLINE=93107072; PubMed=8380170;
RA   Akiyama M., Crooke E., Kornberg A.;
RT   "An exopolyphosphatase of Escherichia coli. The enzyme and its ppx
RT   gene in a polyphosphate operon.";
RL   J. Biol. Chem. 268:633-639(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97349980; PubMed=9205837;
RA   Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA   Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA   Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA   Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,
RA   Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA   Yamagata S., Horiuchi T.;
RT   "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT   - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT   analysis of its sequence features.";
RL   DNA Res. 4:91-113(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O6.H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
RN   [7]
RP   SIMILARITY TO GPPA.
RX   MEDLINE=94025037; PubMed=8212131;
RA   Reizer J., Reizer A., Saier M.H. Jr., Bork B., Sander C.;
RT   "Exopolyphosphate phosphatase and guanosine pentaphosphate
RT   phosphatase belong to the sugar kinase/actin/hsp 70 superfamily.";
RL   Trends Biochem. Sci. 18:247-248(1993).
CC   -!- FUNCTION: DEGRADATION OF INORGANIC POLYPHOSPHATES. ORTHOPHOSPHATE
CC   IS RELEASED PROGRESSIVELY FROM THE ENDS OF POLYPHOSPHATE OF CIRCA
CC   500 RESIDUES LONG, BUT CHAINS OF CIRCA 15 RESIDUES COMPETE POORLY
CC   WITH POLYPHOSPHATE AS SUBSTRATE.
CC   -!- CATALYTIC ACTIVITY: (Polyphosphate)(N) + H(2)O =
CC   (polyphosphate)(N-1) + phosphate.
CC   -!- COFACTOR: Magnesium.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: Membrane-associated.
CC   -!- SIMILARITY: Belongs to the gppA / ppx family.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; L06129; AAA24415.1; -.
DR   EMBL; AE000336; AAC75555.1; -.
DR   EMBL; D90878; BAA16390.1; -.
DR   EMBL; D90880; BAA16392.1; -.
DR   EMBL; AE016764; AAN81470.1; -.
DR   EMBL; AE005479; AAG57612.1; -.
DR   EMBL; AP002561; BAB36787.1; -.
DR   PIR; A45333; A45333.
DR   PIR; D91049; D91049.
DR   PIR; H85893; H85893.
DR   EcoGene; EG11403; ppx.
DR   InterPro; IPR003695; Ppx_GppA.
DR   Pfam; PF02541; Ppx-GppA_1; 1.
KW   Hydrolase; Magnesium; Membrane; Complete proteome.
FT   INIT_MET   0   0
SQ   SEQUENCE   512 AA;   58004 MW;   4861AFF5D9FB9C3 CRC64;

Query Match   7.8%; Score 75.5; DB 1; Length 512;
Best Local Similarity 21.1%; Pred. No. 11;
Matches   48; Conservative   32; Mismatches   79; Indels   69; Gaps   11;

Qy   20 PGGMRGIRGADFQCFQQAR--------------------------AAGLAGTPRA------ 48
              | : ||| ::|:   :|:|| :|:
Db   177 PGGV-----INKENFQRARMAAAQKLETJJTWQFRIQGWNVAMGASGTIKAAHEVLMEMGE 231

Qy   49 ---FLSSRLQDLY-SIVRRADRTGVPVVNLRDE--VLFPSWEALFSG--------------- 89
             ||| ::| | :|: ||:| :| : || ||
Db   232 KDGIITPERLEKLVKEVLRHRNFASLSLPGLSEERKTVFVPGLAILCGVFDALAIRELRL 291

Qy   90 SEGQLKPGARIFSFDGR------DVLQHPAWPRKSVWHGSDPSGRRLTDS---YCETWRTEA 142
             ||: | : : ||| :   :: |:| ||| | | | || |: |:

```
Db      292 SDGALREGV-LYEMEGRFRHQDVRSRTASSLANQYHIDSEQARRVLDTTMQMYEQWREQQ 350

Qy      143 P-AATGQASSLLAGRLLEQEAA------SCRHAFVVLCIENSVMTSFSK 184
                :||||         :      ||||:   ||:
Db      351 PKLAHPQLEALLRWAAMLHEVGLNINHSGLHRHSAYILQNSDLPGFNQ 398


RESULT 13
PIM3_HUMAN
ID   PIM3_HUMAN      STANDARD;      PRT;      326 AA.
AC   Q86V86;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Serine/threonine-protein kinase pim-3 (EC 2.7.1.37).
GN   PIM3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymph;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   IDENTIFICATION FROM ESTS.
RX   MEDLINE=22682943; PubMed=12798037;
RA   Chichester C., Nikitin F., Ravarini J.-C., Lisacek F.;
RT   "Consistency checks for characterizing protein forms.";
RL   Comput. Biol. Chem. 27:29-35(2003).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. PIM
CC       subfamily.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; BC052239; AAH52239.1; ALT_INIT.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_Thr_pkinase.
DR   InterPro; IPR008271; Ser_Thr_pkin_AS.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; Pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation.
FT   DOMAIN       40    293       PROTEIN KINASE.
FT   NP_BIND      46     54       ATP (BY SIMILARITY).
FT   BINDING      69     69       ATP (BY SIMILARITY).
FT   ACT_SITE    170    170       BY SIMILARITY.
SQ   SEQUENCE   326 AA;  35863 MW;  41FDF99DD2467A162 CRC64;

Query Match             7.6%;  Score 73.5;  DB 1;  Length 326;
Best Local Similarity   26.1%;  Pred. No. 9.9;
Matches   42;  Conservative  12;  Mismatches   82;  Indels  25;  Gaps   5;

Qy       12 LVALNSPQPGGMRGIRGADFQCFQQARAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPV 71
               ||:||         |||         ||:
Db      120 LLVLERPEP-------AQDLFDFITERGALDEPLARRFFAQVL---AAVRHCHSCGVVH 168

Qy       72 VNLRDEVLFPSWEA---LFSGSEQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPS- 126
               :::||:|            :||      |||||         :|| :||:
Db      169 RDIKDENLLVDLRSGELKLIDFGSSGALLKDTVYTDFDGTRVYSPPEWIRYHYGRSATV 228

Qy      127 --GRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAAS 164
                 |||||:|  |||     ||||||||||||||||:
Db      229 WSLGVLLYDMVCGDIPFE------QDEEILRGRILFRRRVS 263


RESULT 14
YCDS_ECOLI
ID   YCDS_ECOLI      STANDARD;      PRT;      807 AA.
AC   P75907;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein ycdS precursor.
GN   YCDS OR BI024 OR 21526 OR ECS1270.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562, B3334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
```

RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC   -!- SIMILARITY: STRONG, TO Y.PESTIS HEMIN-BINDING PROTEIN HMSH.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AE000204; AAC74109.1; -.
DR   EMBL; D90739; BAA35806.1; -.
DR   EMBL; D90740; BAA35809.1; -.
DR   EMBL; AE005302; AAG55642.1; -.
DR   EMBL; AP002554; BAB34691.1; -.
DR   PIR; F64844; F64844.
DR   PIR; F90787; F90787.
DR   Ecogene; EG13865; ycdS.
DR   InterPro; IPR008940; Prenyl_trans.
DR   InterPro; IPR011440; TPR.
KW   Hypothetical protein; Outer membrane; Signal; Complete proteome.
FT   SIGNAL   1   26   POTENTIAL.
FT   CHAIN   27   807   HYPOTHETICAL PROTEIN YCDS.
SQ   SEQUENCE   807 AA; 92207 MW; B20067C3D41723FD CRC64;

Query Match   7.6%; Score 73.5; DB 1; Length 807;
Best Local Similarity   26.1%; Pred. No. 28;
Matches   42; Conservative 16; Mismatches   60; Indels   43; Gaps   9;

Qy   2 THQDFQ--LVLHLVALNSPQ-PGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDL 57
          | ||:   :: ::|| | | || || | ||:|:|| |:|:|| | | ||:|
Db   481 TLQEWQQAAVLTHDVVEREFQDPGVVRLKRAVDVHNLABLRIAGSTG------------ 527

Qy   58 YSIVRRADRTGVPVVNLRDEVLFP----SWE--ALFGSSEGQLKPGARIFSPDGRDVLQH 111
          :| |::| |:| | ::| |: | ||: ::|:|| | || |::
Db   528 --IDAEGPDSGKHDVDLTTIVYSPPLKDNWRGFAGFGYADGQFSEGKGIV----RDWLAG 581

Qy   112 PAWPRKSVW-------------HGSDPSGRRLTDSY--CETWR 139
          | : ::| | ||: | : | ||
Db   582 VEWRSRNIWLEAEYAERVFNHEHKP-GARLSGWYDFNDNWR 621

RESULT 15
NA95_HUMAN
ID   NA95_HUMAN   STANDARD;   PRT;   646 AA.
AC   Q9ULX6; O94792; Q96J58; Q9NRQO; Q9UGMO;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Neighbor of A-Kinase anchoring protein 95 (Homologous to AKAP95
DE   protein) (HA95) (Helicase A-binding protein 95) (HAP95) (HRIHFB2018).
GN   NAKAP95.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;

RT   MEDLINE=20163068; PubMed=10697960;
RA   Seki N., Ueki N., Yano K., Saito T., Masuho Y., Muramatsu M.-A.;
RT   "cDNA cloning of a novel human gene NAKAP95, neighbor of A-kinase
RT   anchoring protein 95 (AKAP95) on chromosome 19p13.11-p13.12 region.";
RL   J. Hum. Genet. 45:31-37(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20222332; PubMed=10761695;
RA   Orstavik S., Eide T., Collas P., Han I.O., Taaken K., Kieff E.,
RA   Jahnsen T., Skalhegg B.S.;
RT   "Identification, cloning and characterization of a novel nuclear
RT   protein, HA95, homologous to A-kinase anchoring protein 95.";
RL   Biol. Cell 92:27-37(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta, and T-cell lymphoma;
RX   MEDLINE=20347256; PubMed=10748171;
RA   Westberg C., Yang J.-P., Tang H., Reddy T.R., Wong-Staal F.;
RT   "A novel shuttle protein binds to RNA helicase A and activates the
RT   retroviral constitutive transport element.";
RL   J. Biol. Chem. 275:21396-21401(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SEQUENCE OF 1-358 FROM N.A., AND SUBCELLULAR LOCATION.
RC   TISSUE=Fetal brain;
RX   MEDLINE=99068504; PubMed=9853615;
RA   Ueki N., Oda T., Kondo M., Yano K., Noguchi T., Muramatsu M.-A.;
RT   "Selection system for genes encoding nuclear-targeted proteins.";
RL   Nat. Biotechnol. 16:1338-1342(1998).
CC   -!- FUNCTION: Could play a role in constitutive transport element
CC   (CTE)-mediated gene expression. Does not seem to be implicated in
CC   the binding of regulatory subunit II of PKA.
CC   -!- SUBUNIT: Binds to the C-terminal of RNA helicase A.
CC   -!- SUBCELLULAR LOCATION: Nuclear at steady state but shuttles between
CC   the nucleus and cytoplasm.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AB025905; BAA85003.1; -.
DR   EMBL; AJ243467; CAB55092.1; -.
DR   EMBL; AF199414; AAF86048.1; -.
DR   EMBL; BC000713; AAH00713.1; -.
DR   EMBL; AB015332; BAA34791.1; ALT_INIT.
DR   GO; GO:0005634; C:nucleus; TAS.

```
DR   GO; GO:0017151; F:DEAD/H-box RNA helicase binding; TAS.
DR   InterPro; IPR007071; AKAP95.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF04988; AKAP95; 1.
DR   SMART; SM00355; ZnF_C2H2; 1.
KW   Nuclear protein; Zinc-finger.
FT   DOMAIN      274    279       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      362    364       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   ZN_FING     391    413       C2H2-TYPE.
FT   ZN_FING     484    507       C2H2-TYPE.
FT   DOMAIN       41     52       TYR/GLY-RICH.
FT   DOMAIN      602    608       PRO-RICH.
FT   DOMAIN      589    597       PRO-RICH.
FT   CONFLICT    100    100       D -> N (IN REF. 3).
FT   CONFLICT    189    189       S -> N (IN REF. 3).
FT   CONFLICT    351    358       ALTTQDEN -> EFSWGAWC (IN REF. 4).
FT   CONFLICT    458    458       Q -> H (IN REF. 1).
FT   CONFLICT    554    596       EEEKEQEEAEGGALDEGAQGEAAGISEGAEGVPAQPPVPPE
FT                                PA -> RRRSRRRLRAVPWTRGRRAKRQGFRRAQRACRRS
FT                                LPCPQSQP (IN REF. 3).
FT   CONFLICT    610    640       EEEEGAVPLLGGALQRQIRGIPGLDVEDDEE -> GGGGGR
FT                                RGPCWEGRCNARSAASRASTWRTTRKK (IN REF. 3).
SQ   SEQUENCE   646 AA;  71640 MW;  CBC265BF25996BA2 CRC64;

Query Match             7.5%;  Score 72.5;  DB 1;  Length 646;
Best Local Similarity   26.2%;  Pred. No. 27;
Matches    39;  Conservative   11;  Mismatches   68;  Indels   31;  Gaps    8;

Qy      3 HQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVR  62
          :  :  :|||  ||| : || :: ::  :     :  | :      || |||: :|||
Db     96 NQRLDMVPHLE--TDMMQGGVYGSGGERYDSYESCDS-------RAVLSER--DLY-----  140

Qy     63 RADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHG  122
           |:|:            |     :                   :|   ||
Db    141 --RSGYDYSELD------PEMEMAY---EGQYDAYRDQFRMRGNDTF----GPRAQGWAR  185

Qy    123 SDPSGRRLTDSYCETWRTEAPAATGQASS  151
             |||   :       | :  |
Db    186 DARSGRPMASGYGRMWE-DPMGARGQCMS  213


Search completed: March 26, 2004, 13:36:34
Job time : 12.3333 secs
```

OM protein - protein search, using sw model

```
                  GenCore version 5.1.6
            Copyright (c) 1993 - 2004 Compugen Ltd.
```

Run on: March 26, 2004, 13:12:37 ; Search time 32.4444 Seconds
                                         (without alignments)
                                         1789.377 Million cell updates/sec

Title: US-09-938-391-4
Perfect score: 966
Sequence: 1 HTHQDFQLVLHLVALNSPQP.........CRHAFVVLCIENSVMTSFSK 184

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SPTREMBL_25:*
                1: sp_archea:*
                2: sp_bacteria:*
                3: sp_fungi:*
                4: sp_human:*
                5: sp_invertebrate:*
                6: sp_mammal:*
                7: sp_mhc:*
                8: sp_organelle:*
                9: sp_phage:*
               10: sp_plant:*
               11: sp_rodent:*
               12: sp_virus:*
               13: sp_vertebrate:*
               14: sp_unclassified:*
               15: sp_rvirus:*
               16: sp_bacteriap:*
               17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Query Length	DB	ID	Description
1	835	86.4	187	4	Q8WXI5	Q8wxi5 homo sapien
2	835	86.4	261	4	Q8NG19	Q8ng19 homo sapien
3	835	86.4	816	4	Q8N4S4	Q8n4s4 homo sapien
4	831	86.0	1140	11	Q61434	Q61434 mus musculu
5	813	84.2	226	11	Q9QZD2	Q9qzd2 rattus norv
6	754	78.1	171	11	Q9WTW5	Q9wuw5 rattus norv
7	750	77.6	1344	13	Q93419	Q93419 gallus gall
8	726	75.2	160	11	Q9CRT2	Q9crt2 mus musculu
9	694	71.8	1315	13	Q8QHL9	Q8qhl9 xenopus lae
10	691	71.5	1307	13	Q8JFF7	Q8jff7 xenopus lae
11	631.5	65.4	361	13	Q8AWC6	Q8awc6 brachydanio
12	551	57.0	1367	11	O35206	O35206 mus musculu
13	550	56.9	1367	11	Q9EQD9	Q9eqd9 mus musculu
14	527	54.6	1388	4	Q9Y4W4	Q9y4w4 homo sapien
15	457.5	47.4	950	5	Q86SC8	Q86sc8 ciona intes
16	456	47.2	102	4	Q96T70	Q96t70 homo sapien
17	434.5	45.0	299	5	Q8MSE3	Q8mse3 drosophila
18	429.5	44.5	792	5	Q8MT89	Q8mt89 drosophila
19	428.5	44.4	778	5	Q86BH1	Q86bh1 drosophila
20	366	37.9	650	5	O17866	O17866 caenorhabdi
21	366	37.9	778	5	Q9U9K6	Q9u9k6 caenorhabdi
22	366	37.9	1117	5	Q9U9K7	Q9u9k7 caenorhabdi
23	366	37.9	1154	5	Q8I0G6	Q8i0g6 caenorhabdi
24	101.5	10.5	287	16	Q89I38	Q89i38 bradyrhizob
25	94.5	9.8	208	16	Q92K28	Q92k28 rhizobium m
26	93.5	9.7	702	2	O52660	O52660 escherichia
27	93.5	9.7	713	2	O52679	O52679 escherichia
28	87	9.0	651	2	Q9VFA9	Q9vfa9 drosophila
29	85.5	8.9	6521	2	Q8RL72	Q8rl72 pseudomonas
30	84.5	8.7	282	16	Q8PJG2	Q8pjg2 xanthomonas
31	84	8.7	539	2	Q914Y1	Q914y1 streptomyce
32	83.5	8.6	240	16	Q8F863	Q8f863 leptospira
33	83	8.6	557	16	Q9K4L0	Q9k4l0 streptomyce
34	82	8.5	687	4	Q86TP6	Q86tp6 homo sapien
35	82	8.5	891	2	O33468	O33468 pseudomonas
36	82	8.5	1050	16	Q8PAG4	Q8pag4 xanthomonas
37	82	8.5	1105	4	Q86XT1	Q86xt1 homo sapien
38	82	8.5	1113	4	Q86WQ3	Q86wq3 homo sapien
39	81.5	8.4	633	16	Q9RIZ7	Q9riz7 streptomyce
40	81.5	8.4	633	16	Q8XBY5	Q8xby5 escherichia
41	80.5	8.3	314	16	Q82UJ4	Q82uj4 nitrosomona
42	80.5	8.3	569	5	Q9N6F0	Q9n6f0 leishmania
43	79.5	8.2	276	16	Q89P16	Q89p16 bradyrhizob
44	79.5	8.2	346	16	Q9HWN2	Q9hwn2 pseudomonas
45	79.5	8.2	713	16	Q8XEE0	Q8xee0 escherichia

ALIGNMENTS

RESULT 1
Q8WXI5
```
ID   Q8WXI5;        PRELIMINARY;       PRT;     187 AA.
AC   Q8WXI5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Collagen XVIII (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21409408; PubMed=11517600;
RA   Feng Y., Cui L.B., Liu C.X., Ma Q.J.;
RT   "Inhibition effect in vitro of purified endostatin expressed in Pichia
RT   pastoris.";
RL   Sheng Wu Gong Cheng Xue Bao 17:278-282(2001).
DR   EMBL; AF416592; AAL37720.1; -.
FT   NON_TER       1
SQ   SEQUENCE   187 AA;  20448 MW;  72B1047D85838CD3 CRC64;

Query Match            86.4%;  Score 835;  DB 4;  Length 187;
Best Local Similarity  85.1%;  Pred. No. 5.1e-76;
Matches  154;  Conservative  14;  Mismatches  13;  Indels  0;  Gaps  0;

Qy    1 HTHQDFQLVLHLVALNSPQPGGMGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI  60
         |:::||  |||||||||||   ||||||||||||||||| ||||||||||||||||||||
Db    5 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI  64

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
         |||||:  :|:|||:|||:|||:|||||||||:|||||||||||:||| ::|| || :||
Db   65 VRRADRAAVPIVNLKDELLPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW  124

Qy  121 HGSDPSGRRLLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
         |||||.||||| |||:|||||||.||||||||||.||||.|.|||| ||::|||||||.||
Db  125 HGSDPNGRRLLTESYCETWRTEAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 184
```

Left column

```
Qy   181 S 181
          :
Db   185 A 185

RESULT 2
Q8NG19
ID   Q8NG19_      PRELIMINARY;      PRT;   261 AA.
AC   Q8NG19;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Multi-functional protein MFP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Dou D.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF282883; AAM52249.1; -.
SQ   SEQUENCE 261 AA;  26745 MW;  CA60C920AF3E90S5 CRC64;

Query Match         86.4%;  Score 835;  DB 4;  Length 261;
Best Local Similarity 85.1%;  Pred. No. 7.8e-75;
Matches 154;  Conservative 14;  Mismatches 13;  Indels 0;  Gaps 0;

Qy    1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
          |:||||:|||||||||||||:|||||||||||||||||:|||||||||||||||||||
Db   79 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 138

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGLKPGARIFSFDGRDVLQHPAWPRKSVW 120
          |||||||:|||||:|||||||||||||||||||||||||||||||||||||:|:|||
Db  139 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 198

Qy  121 HGSDPSGRRLTDSYCETWRTEAPAATQOASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
          ||||||||||||||||||||||||||||||||||||||||||::||:|||||||||||
Db  199 HGSDPNGRRLTESYCETWRTEAPSATQOASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 258

Qy  181 S 181
          :
Db  259 A 259

RESULT 3
Q8N4S4
ID   Q8N4S4_      PRELIMINARY;      PRT;   816 AA.
AC   Q8N4S4;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Similar to collagen, type XVIII, alpha 1 (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Renal adenocarcinoma;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC033715; AAH33715.1; -.
DR   InterPro; IPR008161; Clg.helix.
DR   InterPro; IPR008160; Collagen.
DR   Pfam; PF01391; Collagen; 5.
DR   ProDom; PD000007; Clg_helix; 1.
KW   Collagen.
FT   NON_TER       1       1
SQ   SEQUENCE 816 AA;  82553 MW;  5D53B2946694F86 CRC64;

Query Match         86.4%;  Score 835;  DB 4;  Length 816;
Best Local Similarity 85.1%;  Pred. No. 3.3e-75;
```

Right column

```
Matches 154;  Conservative 14;  Mismatches 13;  Indels 0;  Gaps 0;

Qy    1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
          |:||||:|||||||||||||:|||||||||||||||||:|||||||||||||||||||
Db  634 HSHRDFQPVLHLVALNSPLSGGMRGIRGADFQCFQQARAVGLAGTFRAFLSSRLQDLYSI 693

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGLKPGARIFSFDGRDVLQHPAWPRKSVW 120
          |||||||:|||||:|||||||||||||||||||||||||||||||||||||:|:|||
Db  694 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGPLKPGARIFSFDGKDVLRHPTWPQKSVW 753

Qy  121 HGSDPSGRRLTDSYCETWRTEAPAATQOASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
          ||||||||||||||||||||||||||||||||||||||||||::||:|||||||||||
Db  754 HGSDPNGRRLTESYCETWRTEAPSATQOASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 813

Qy  181 S 181
          :
Db  814 A 814

RESULT 4
Q61434
ID   Q61434_      PRELIMINARY;      PRT;   1140 AA.
AC   Q61434;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Collagen (Fragment).
GN   COL15A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Abe N., Muragaki Y., Yoshioka H., Inoue H., Ninomiya Y.;
RT   "Identification of a novel collagen chain represented by extensive
RT   interruptions in the triple-helical region.";
RL   Cell. Mol. Biol. Res. 196:576-582(1993).
DR   EMBL; D17546; BAA04483.1; -.
DR   PIR; B56101; B56101.
DR   HSSP; P39061; 1KOE.
DR   MGD; MGI:88449; Col15a1.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 8.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
KW   Collagen.
FT   NON_TER       1       1
SQ   SEQUENCE 1140 AA;  115156 MW;  8B0C7E6862B3BDFE CRC64;

Query Match         86.0%;  Score 831;  DB 11;  Length 1140;
Best Local Similarity 83.7%;  Pred. No. 1.3e-74;
Matches 154;  Conservative 17;  Mismatches 13;  Indels 0;  Gaps 0;

Qy    1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
          |:||||:|||||||||||||:|||||||||||||||||:|||||||||||||||||||
Db  957 HTHQDFQPVLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSI 1016

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGLKPGARIFSFDGRDVLQHPAWPRKSVW 120
          |||||:|:|||||:|||||::||||||||||:||||||||||||||:|||:|:|||
Db 1017 VRRADRGSVPIVNLKDEVLSPSWDSLFSGSGQLQPGARIFSFDGRDVLRHPAWPQKSVW 1076

Qy  121 HGSDPSGRRLTDSYCETWRTEAPAATQOASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
          ||||||||||:||||||||||||||||||||||||||||:||:||::|:::|||||||||
Db 1077 HGSDPSGRRLMESYCETWRTETTGATQOASSLLSGRLLEQKXAASCHNSYIVLCIENSFMT 1136

Qy  181 SFSK 184
          ||||
Db 1137 SFSK 1140
```

```
RESULT 5
Q9QZD2
ID   Q9QZD2        PRELIMINARY;     PRT;   226 AA.
AC   Q9QZD2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update).
DE   Collagen XVIII (Fragment).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=0227226; PubMed=10766159;
RA   Perletti G., Concari P., Giardini R., Marras E., Piccinini F.,
RA   Folkman J., Chen L.;
RT   "Antitumor activity of endostatin against carcinogen-induced rat
RT   primary mammary tumors.";
RL   Cancer Res. 60:1793-1796(2000).
DR   EMBL; AF189709; AAF00975.1; -.
DR   HSSP; P39061; 1KOE.
FT   NON_TER        1      1
SQ   SEQUENCE   226 AA;   25350 MW;   38B83C0486C0E949 CRC64;

Query Match    84.2%;  Score 813;  DB 11;  Length 226;
Best Local Similarity  82.1%;  Pred. No. 1.1e-73;
Matches 151;  Conservative  16;  Mismatches  17;  Indels    0;  Gaps    0;

Qy    1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
        |||||| |:|||||||| :|||||||||||||||||||||:|:|||||||||||||||||
Db   43 HTHQDFHPVLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSI 102

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSBGGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
        ||||:: |:|:||||:|||| |||: ||||| |:|:|||||||||||||||:||||:|:||
Db  103 VRRADRSSVPIVNLKDEVLSPSWDTLFSGSGQLHSGARIFSFDGRDVLRHPAWPQKSVW 162

Qy  121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
        |||||||||||:|||||||||| :|||||||||:|||||||:||| : ::|||||||:||
Db  163 HGSDPSGRRLMESYCETWRTEATGVTGQASSLLSGRLLEQKAESCHNSYIVLCIENSFMT 222

Qy  181 SFSK 184
        ||||
Db  223 SFSK 226


RESULT 6
Q9WUW5
ID   Q9WUW5        PRELIMINARY;     PRT;   171 AA.
AC   Q9WUW5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update).
DE   Collagen type XVIII, alpha (I) chain (Fragment).
GN   COL18A1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Jia J.D., Bauer M., Eberspaecher U., Donner P., Schuppan D.;
RT   "Partial 3'-terminal cDNA sequence of rat collagen XVIII/endostatin.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Jia J.D., Bauer M., Sedlaczek N., Ruehl M., Riecken E.O., Schuppan D.;
RT   "Temporospatial expression of collagen XVIII/endostatin in acute and
RT   chronic liver injuries.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ236873; CAB44263.1; -.
DR   HSSP; P39061; 1KOE.
FT   NON_TER        1      1
FT   NON_TER      171    171
SQ   SEQUENCE   171 AA;   18933 MW;   81BE2E83FC2C8B72 CRC64;

Query Match    78.1%;  Score 754;  DB 11;  Length 171;
Best Local Similarity  81.9%;  Pred. No. 6.5e-68;
Matches 140;  Conservative  16;  Mismatches  15;  Indels    0;  Gaps    0;

Qy    9 VLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTG 68
        |||||||:|:|||||||||||||||||||||||||||||||||||||||||:|||||:||
Db    1 VLHLVALNTPLSGGMRGIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRSS 60

Qy   69 VPVVNLRDEVLFPSWEALFSGSEGGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGR 128
        |:||||:|||| |||: |||||| |:|:|||||||||||||||:||||:|:||||||||||
Db   61 VPIVNLKDEVLSPSWDTLFSGSGQLHSGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGR 120

Qy  129 RLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVM 179
        ||:|||||||||||| :|||||||||||||||:|||| : ::|||||||||
Db  121 RLMESYCETWRTEATGVTGQASSLLSGRLLEQKAESCHNSYIVLCIENSFM 171


RESULT 7
O93419
ID   O93419        PRELIMINARY;     PRT;   1344 AA.
AC   O93419;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Collagen XVIII precursor.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98411346; PubMed=9738008;
RA   Halfter W., Dong S., Schurer B., Cole G.J.;
RT   "Collagen XVIII is a basement membrane heparan sulfate proteoglycan.";
RL   J. Biol. Chem. 273:25404-25412(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Halfter W., Dong S.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF083440; AAC33294.2; -.
DR   HSSP; P39061; 1KOE.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008161; C1q.helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008995; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 8.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; C1q_helix; 2.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen; Signal.
FT   SIGNAL         1     26        POTENTIAL.
SQ   SEQUENCE   1344 AA;   137402 MW;   7AA366B4FB940CCD CRC64;

Query Match    77.6%;  Score 750;  DB 13;  Length 1344;
Best Local Similarity  76.1%;  Pred. No. 2.3e-66;
Matches 140;  Conservative  16;  Mismatches  28;  Indels    0;  Gaps    0;

Qy    1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
        ||||||:|||||||||||:|||||||||||||||||||||:|:||||||||||||||||||
Db 1161 HVHQDFQPALHLVALNTPLSGGMRGIRGADFQCFQQARQVGLAGTFRAFLSSRLQDLYSI 1220
```

```
Qy     61 VRRADRTGVPVVNLRDEVLFPSWEALFSSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
          ||||||: ||:|||||||||||: :||||||:    |: ||:||||||||||: :|||||
Db   1221 VRRADRTAVPIVNLRDEVLFPSNWEALFTGSEAPLRAGARILSFDGRDILQDSAWPQKSIW 1280

Qy    121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
          |||| :|||| :|||:|||||: ||||||||| :|:||||: ||||:||||||||||:||
Db   1281 HGSDAKGRRLPESYCEAWRTDERGTSGQASSLSSGKLLEQSASSCQHAFVVLCIENSFMT 1340

Qy    181 SFSK 184
          : ||
Db   1341 AAKK 1344
```

RESULT 8
Q9CRT2
```
ID   Q9CRT2      PRELIMINARY;      PRT;   160 AA.
AC   Q9CRT2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Procollagen, type XVIII, alpha 1 (Fragment).
GN   COL18A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki N., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang J., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK014292; BAB29249.1; -.
DR   HSSP; P39061; 1KOE.
DR   MGD; MGI:88451; Col18a1.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   GO; GO:0001525; P:angiogenesis; IMP.
FT   NON_TER       1      1
SQ   SEQUENCE   160 AA;  17725 MW;  60F8S3D777C375D2 CRC64;
```

```
Query Match             75.2%;  Score 726;  DB 11;  Length 160;
Best Local Similarity   83.8%;  Pred. No. 4e-65;
Matches  134;  Conservative  16;  Mismatches  10;  Indels  0;  Gaps  0;

Qy     25 GIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGVPVVNLRDEVLFPSWE 84
          ||||||||||||||| ||:|||||||||||||||||||||||: |:|||||||||| |||
Db      1 GIRGADFQCFQQARAVGLSGTFRAFLSSRLQDLYSIVRRADRGSVPIVNLRDEVLSPSWD 60

Qy     85 ALFSSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRRLTDSYCETWRTEAPA 144
          :|||: | |||:|||||||||||||||:||||:|||||||||||||||::|||||||| |
Db     61 SLFSGSGQGQLQPGARIFSFDGRDVLRHPAWPQKSVWHGSDPSGRRLMESYCETWRTETTG 120

Qy    145 ATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 184
          ||||||||| ||||||: |||:| :|||||||||| |||
```

```
Db    121 ATGQASSLLGRLLEQKAASCHNSYIVLCIENSFMTSFSK 160
```

RESULT 9
Q8QHL9
```
ID   Q8QHL9      PRELIMINARY;      PRT;   1315 AA.
AC   Q8QHL9;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Type XVIII collagen alpha1 chain.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ishino T., Sekimizu K., Natori S., Kubo T.;
RT   "Identification and characterization of genes expressed selectively in
RT   the regenerating tail of Xenopus laevis tadpole.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB047066; BAB84674.1; -.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 7.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen.
SQ   SEQUENCE   1315 AA;  134946 MW;  0C56C235DE058355 CRC64;
```

```
Query Match             71.8%;  Score 694;  DB 13;  Length 1315;
Best Local Similarity   70.1%;  Pred. No. 9.6e-61;
Matches  129;  Conservative  18;  Mismatches  37;  Indels  0;  Gaps  0;

Qy      1 HTHQDFQLVLHLVALNSPQFGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
          ||||:|:  |:|||||:| ||:::| |||||||| ||||:|:||||||||||||||||||
Db   1132 HTHQQFNPALHIVALNAPLSGSMKSIRGVDFQCFEQQARKSGLHGTFRAFLSSRLQDLYSI 1191

Qy     61 VRRADRTGVPVVNLRDEVLFPSWEALFSSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
          ||||||: : :|||||||| :.|| :||||:|  :|||||||||||:|| |:| || |:||
Db   1192 VRRADRQSVQIVNLRDEVLYDNWESLFSGSEAQMRPGARIFSFDGKDVATDPTWPQKMVW 1251

Qy    121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
          ||||: |||| :|||||||||: | ||||||||: |:||||: ||| :|:||||||||:||
Db   1252 HGSDAKGRRLTESYCETWRTDESAVTSQASSLTSGKLLEQRPQSCNKNFIVLCIENSFMT 1311

Qy    181 SFSK 184
          : :|
Db   1312 NNRK 1315
```

RESULT 10
Q8JFF7
```
ID   Q8JFF7      PRELIMINARY;      PRT;   1307 AA.
AC   Q8JFF7;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Type XVIII collagen short variant.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22166979; PubMed=12175494;
RA   Elamaa H., Peterson J., Pihlajaniemi T., Destree O.;
```

RT   "Cloning of three variants of type XVIII collagen and their expression
RT   patterns during Xenopus laevis development.";
RL   Mech. Dev. 114:109-113(2002).
DR   EMBL; AY052763; AAL14257.1; -.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 6.
DR   Pfam; PF02210; TSPN; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen.
SQ   SEQUENCE   1307 AA;   134153 MW;   D53EDBFE3DE34976 CRC64;

Query Match   71.5%; Score 691; DB 13; Length 1307;
Best Local Similarity   70.1%; Pred. No. 1.9e-60;
Matches 129; Conservative 18; Mismatches 37; Indels 0; Gaps 0;

Qy   1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
          |||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||:
Db   1124 HTHQDFNPALHLVALNAPISGSMKSIRGVDFQCFEQARKAGLHGTFRAFLSSRLQDLYSI 1183

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
          |||||| ||||||||||||||:|:||||||||:||||||:|||||||::|:||
Db   1184 VRRADRQSVQIVNLRDEVLYDNWDSLFSGSSEAQMRSGARILSFDGKDVTTDFTWPQKVVW 1243

Qy   121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
          ||||||| |||||||||||||||||||||::||||||||||||:||||:||||||||||
Db   1244 HGSDAKGRRLTESYCETWRTDESAVTGQASSLTSGKLLEQRPQSCNKNFIVLCIENSFMT 1303

Qy   181 SFSK 184
          ||||
Db   1304 NNRK 1307

RESULT 11
Q8AWC6
ID   Q8AWC6;   PRELIMINARY;   PRT;   361 AA.
AC   Q8AWC6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Collagen XVIII (Fragment).
GN   COL18A1.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Haftek Z., Morvan-Dubois G., Thisse B., Garrone R., Le Guellec D.;
RT   "Sequence and embryonic expression of collagen XVIII NC11 domain
RT   (endostatin) in the zebrafish.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ494837; CAD38825.1; -.
FT   NON_TER   1   1
SQ   SEQUENCE   361 AA;   40222 MW;   3C5A0F8479D26735 CRC64;

Query Match   65.4%; Score 631.5; DB 13; Length 361;
Best Local Similarity   65.2%; Pred. No. 3.7e-55;
Matches 120; Conservative 22; Mismatches 41; Indels 1; Gaps 1;

Qy   1 HTHQDFQLVLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSI 60
          |||||||||||||||||:|||:||||||||||||||||||||||||||||||||||||||
Db   179 HTHTSGP-GLHIJALNSPQVGNMRGIRGADFMCFQQARAVGMKGTFRAFLSSRLQDLYSI 237

Qy   61 VRRADRTGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVW 120
          ||||||| ||||||||||||||:|||||||||:|||||||||||||||:|||||||||
Db   238 VRRSDRETLPIVNLKDQVLFSSWESLFSDSESRMKDNAPIIYSFDGRDVLRDSANPEKMIW 297

Qy   121 HGSDPSGRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT 180
          |||||| |:|||||||||| |||||:|:|||||:||
Db   ...

Db   298 HGSDCRGHRQTDNYCETWRAGDRAVTGLASSLQAGQLLQQTSSSCSSSYIALCIENSVMT 357

Qy   181 SFSK 184
          |
Db   358 QSKK 361

RESULT 12
O35206
ID   O35206;   PRELIMINARY;   PRT;   1367 AA.
AC   O35206;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type XV collagen.
GN   COL15A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9480713; PubMed=9339358;
RA   Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA   Pihlajaniemi T.;
RT   "Cloning of mouse type XV collagen sequences and mapping of the
RT   corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
RT   (XV) collagen sequences indicates divergence in the number of small
RT   collagenous domains.";
RL   Genomics 45:31-41(1997).
DR   EMBL; AF011450; AAC53387.1; -.
DR   PDB; 1DY2; 21-JAN-01.
DR   MGD; MGI:88449; Col15a1.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 5.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen.
SQ   SEQUENCE   1367 AA;   140525 MW;   A483A1254AF3ABEC CRC64;

Query Match   57.0%; Score 551; DB 11; Length 1367;
Best Local Similarity   59.7%; Pred. No. 2.5e-46;
Matches 105; Conservative 25; Mismatches 42; Indels 4; Gaps 1;

Qy   9 VLHLVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTG 68
          ||||||||||||||||:|||||||:|||| ::|||||||||||||||||||:||:|:|
Db   1196 VLHLVALNTPVAGDIR---ADFQCFQQARAAGLLSTFRAFLSSHLQDLSTVVRKAERFG 1251

Qy   69 VPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGR 128
          :|||||||||||:||||:||||||:|:|||||||||||||:||||||:|||||||||||
Db   1252 LPIVNLKGQVLFNNWDSIFSGDGGQPNTHIPIYSFDGRDVMTDFSWPQKVVWHGSNPHGV 1311

Qy   129 RLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 184
          ||||||||||||:|||||:|||||||:|:||||| :||:||||||||||||:|
Db   1312 RLVDKYCEAWRTDMAVTGFASPLSTGKILDQKAYSCANRLIVLCIENSFMTDTRK 1367

RESULT 13
Q9EQD9
ID   Q9EQD9;   PRELIMINARY;   PRT;   1367 AA.
AC   Q9EQD9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type XV collagen.

Left column

```
GN   COL15A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=97480713; PubMed=9333358;
RA   Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA   Pihlajaniemi T.;
RT   "Cloning of mouse type XV collagen sequences and mapping of the
RT   corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
RT   (XV) collagen sequences indicates divergence in the number of small
RT   collagenous domains.";
RL   Genomics 45:31-41(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=20522048; PubMed=11068203;
RA   Eklund L., Muona A., Lietard J., Pihlajaniemi T.;
RT   "Structure of the mouse type XV collagen gene, Col15a1, comparison
RT   with the human COL15A1 gene and functional analysis of the promoters
RT   of both genes.";
RL   Matrix Biol. 19:489-500(2000).
DR   EMBL; AF261131; AAG27545.1; -.
DR   EMBL; AF261109; AAG27545.1; JOINED.
DR   EMBL; AF261110; AAG27545.1; JOINED.
DR   EMBL; AF261111; AAG27545.1; JOINED.
DR   EMBL; AF261112; AAG27545.1; JOINED.
DR   EMBL; AF261113; AAG27545.1; JOINED.
DR   EMBL; AF261114; AAG27545.1; JOINED.
DR   EMBL; AF261115; AAG27545.1; JOINED.
DR   EMBL; AF261116; AAG27545.1; JOINED.
DR   EMBL; AF261117; AAG27545.1; JOINED.
DR   EMBL; AF261118; AAG27545.1; JOINED.
DR   EMBL; AF261119; AAG27545.1; JOINED.
DR   EMBL; AF261120; AAG27545.1; JOINED.
DR   EMBL; AF261121; AAG27545.1; JOINED.
DR   EMBL; AF261122; AAG27545.1; JOINED.
DR   EMBL; AF261123; AAG27545.1; JOINED.
DR   EMBL; AF261124; AAG27545.1; JOINED.
DR   EMBL; AF261125; AAG27545.1; JOINED.
DR   EMBL; AF261126; AAG27545.1; JOINED.
DR   EMBL; AF261127; AAG27545.1; JOINED.
DR   EMBL; AF261128; AAG27545.1; JOINED.
DR   EMBL; AF261129; AAG27545.1; JOINED.
DR   EMBL; AF261130; AAG27545.1; JOINED.
DR   HSSP; P39061; 1KOE.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 5.
DR   Pfam; PF02210; TSPN; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
KW   Collagen.
SQ   SEQUENCE   1367 AA;  140454 MW;  40F259A3CD47C982 CRC64;

Query Match             56.9%;  Score 550;  DB 11;  Length 1367;
Best Local Similarity   59.7%;  Pred. No. 3.2e-46;
Matches  105;  Conservative  25;  Mismatches  42;  Indels   4;  Gaps   1;

Qy     9 VLHLVALNSPQPGGMRGIRGADPQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTG 68
           |||||||:|:| ||| ||:||||||||||||||| |||||||:|||| ::||:|:|
Db  1196 VLHLVALNTPVAGDIR----ADPQCFQQARAAGLLSTFRAFLSSHLQDLSTVVRKAERFG 1251

Qy    69 VPVVNLRDEVLFPSWEALFSGSSGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGR 128
```

Right column

```
Db  1252 LPIVNLKGQVLFNWWDSIFSGDGGQFNTHIPIYSFPGRDVMTDPSWPQKVVWHGSNPYGV 1311
           :|:||:|. :|||. :.|||| .: :||  |||||| . : |:.|||:||.| ||||||:| |
Qy   129 RLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK 184
           ||::||.|.:|||| :|||:|| :: | :| | |:|:|.||:|:|:|||:||:| |
Db  1312 RLVDKYCEAWRTTDMAVTGFASPLSTGKILDQKAYSCANRLIVLCIENSFMTDTRK 1367

RESULT 14
Q9Y4W4
ID   Q9Y4W4        PRELIMINARY;        PRT;   1388 AA.
AC   Q9Y4W4;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Type XV collagen.
GN   COL15A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94148920; PubMed=8106446;
RA   Kivirikko S., Heinamaki P., Rehn M., Honkanen N., Myers J.C.,
RA   Pihlajaniemi T.;
RT   "Primary structure of the alpha 1 chain of human type XV collagen and
RT   exon-intron organization in the 3' region of the corresponding gene.";
RL   J. Biol. Chem. 269:4773-4779(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98316357; PubMed=9651385;
RA   Hagg P.M., Muona A., Lietard J., Kivirikko S., Pihlajaniemi T.;
RT   "Complete exon-intron organization of the human gene for the alpha1
RT   chain of type XV collagen (COL15A1) and comparison with the homologous
RT   COL18A1 gene.";
RL   J. Biol. Chem. 273:17824-17831(1998).
DR   EMBL; L25280; AAC78500.1; -.
DR   EMBL; AF052956; AAC78500.1; JOINED.
DR   EMBL; AF052957; AAC78500.1; JOINED.
DR   EMBL; AF052958; AAC78500.1; JOINED.
DR   EMBL; AF052959; AAC78500.1; JOINED.
DR   EMBL; AF052960; AAC78500.1; JOINED.
DR   EMBL; AF052961; AAC78500.1; JOINED.
DR   EMBL; AF052962; AAC78500.1; JOINED.
DR   EMBL; AF052963; AAC78500.1; JOINED.
DR   EMBL; AF052964; AAC78500.1; JOINED.
DR   EMBL; AF052965; AAC78500.1; JOINED.
DR   EMBL; AF052966; AAC78500.1; JOINED.
DR   EMBL; AF052967; AAC78500.1; JOINED.
DR   EMBL; AF052968; AAC78500.1; JOINED.
DR   EMBL; AF052969; AAC78500.1; JOINED.
DR   EMBL; AF052970; AAC78500.1; JOINED.
DR   EMBL; AF052971; AAC78500.1; JOINED.
DR   EMBL; AF052972; AAC78500.1; JOINED.
DR   EMBL; AF052973; AAC78500.1; JOINED.
DR   EMBL; AF052974; AAC78500.1; JOINED.
DR   EMBL; AF052975; AAC78500.1; JOINED.
DR   EMBL; L25285; AAC78500.1; JOINED.
DR   EMBL; L25284; AAC78500.1; JOINED.
DR   EMBL; L25283; AAC78500.1; JOINED.
DR   EMBL; L25282; AAC78500.1; JOINED.
DR   EMBL; L25281; AAC78500.1; JOINED.
DR   HSSP; P39061; 1KOE.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0007155; P:cell adhesion; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   Pfam; PF01391; Collagen; 5.
DR   Pfam; PF02210; TSPN; 1.
DR   SMART; SM00282; LamG; 1.
```

```
DR   SMART; SM00210; TSPN; 1.
KW   Collagen.
SQ   SEQUENCE   1388 AA;   141757 MW;  96828E45E847194B CRC64;

Query Match.
Best Local Similarity   54.6%;   Score 527; DB 4;   Length 1388;
                        56.6%;   Pred. No. 6.8e-44;
Matches   99;  Conservative   27;  Mismatches   45;  Indels   4;  Gaps   1;

Qy      10  LHIVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADRTGV  69
                ||||||| |||:|||  ||||:||||||||:||||||||||| |||:||||:|:|
Db    1218  LHLAALNMPVSGDIR----ADFQCFKQARAAGLLSTYRAFLSSHLQDLSTIVRKAERYSL  1273

Qy      70  PVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPSGRR  129
                :|||:|||:||| :||::|||||  ||||||||||||:  :|:|:||||||||||| |
Db    1274  PIVNLKGQVLFNWDSIFSGHGGQFNMHIPIYSFDGRDIMTDPSWPQKVIWHGSSPHGVR  1333

Qy     130  LTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMTSFSK  184
                :||:||||||||||||||||||||||:||||| :| |||||:||||||  ||||
Db    1334  LVDNYCEAWRTADTAVTGLASPLSTGKILDQKAYSCANRLIVLCIENSFMTDARK  1388

RESULT 15
Q86SC8
ID   Q86SC8        PRELIMINARY;         PRT;    950 AA.
AC   Q86SC8;
DT   01-JUN-2003  (TrEMBLrel. 24, Created)
DT   01-JUN-2003  (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003  (TrEMBLrel. 25, Last annotation update)
DE   Collagen XVIII homologue.
GN   CI-COLLAGEN XVIII.
OS   Ciona intestinalis.
OC   Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC   Phlebobranchia; Cionidae; Ciona.
OX   NCBI_TaxID=7719;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hotta K., Takahashi H., Satoh N.;
RT   "Further Characterization of Brachyury-Downstream Genes in Ciona
RT   intestinalis Embryo.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB076899; BAC57521.1; -
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001759; Pentaxin.
DR   Pfam; PF01391; Collagen; 8.
DR   PROSITE; PS00289; PENTAXIN; 1.
SQ   SEQUENCE   950 AA;   94768 MW;  2767ECA158CBA89F CRC64;

Query Match
Best Local Similarity   47.4%;   Score 457.5; DB 5;   Length 950;
                        55.2%;   Pred. No. 4.2e-37;
Matches   96;  Conservative   17;  Mismatches   60;  Indels   1;  Gaps   1;

Qy       7  QLVLHIVALNSPQPGGMRGIRGADFQCFQQARAAGLAGTFRAFLSSRLQDLYSIVRRADR  66
                ||||||||||||||||||||||||||||||||||||||||| |||:||||||||||
Db     773  QKSLHMIALNFPLRGNTRGIVGADARCFQQARAGLKGTYRAFLSSRDQNVRSIVRREDR  832

Qy      67  TGVPVVNLRDEVLFPSWEALFSGSEGQLKPGARIFSFDGRDVLQHPAWPRKSVWHGSDPS  126
                ||:|:|||||:  |||:: :  :|| |:   |||||||||||:|||||:|| ||||||
Db     833  RNVPIVNIRGQLFSSWEELFR-TEGRMDNPNMIYSFENRQVSTDARWPVKFVWHGSYTD  891

Qy     127  GRRLTDSYCETWRTEAPAATGQASSLLAGRLLEQEAASCRHAFVVLCIENSVMT  180
                ||| :|:|||||||  |:|||| |:| ||||:|||| :| ||||||||||||
Db     892  GRLNPMHYCASWYTDHKAVTGQASPLSTRRLLAQKPYSCESGFVVLCVENSTRT  945
```

Search completed: March 26, 2004, 13:38:00
Job time : 34.4444 secs